

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10404 A2

(51) International Patent Classification⁷: **C12N 15/57**,
15/62, 5/10, 9/64, C12Q 1/68, G01N 33/68, 33/577, A61K
31/711, 31/713, 38/43, 39/395

(21) International Application Number: PCT/EP01/08181

(22) International Filing Date: 16 July 2001 (16.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,211 27 July 2000 (27.07.2000) US

(71) Applicant (for all designated States except US): **BAYER
AKTIENGESELLSCHAFT** [DE/DE]; 51368 Lev-
erkusen (DE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **XIAO, Yonghong**
[US/US]; 78 Dana Street #1, Cambridge, MA 02138 (US).

(74) Common Representative: **BAYER AKTIENGE-
SELLSCHAFT**; 51368 Leverkusen (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished
upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: REGULATION OF HUMAN MASP-LIKE SERINE PROTEASE

(57) Abstract: Reagents which regulate human MASP-like serine protease activity and reagents which bind to human MASP-like serine protease gene products can be used to regulate extracellular matrix degradation. Such regulation is particularly useful for defending against pathogens, such as viruses, bacteria, mycoplasma, fungi, protozoa, helminths, rickettsia, chloamydiae, parasites, prions, and the like.

WO 02/10404 A2

A4 - 09/981,123

REGULATION OF HUMAN MASP-LIKE SERINE PROTEASE

TECHNICAL FIELD OF THE INVENTION

The invention relates to the area of regulation of MASP-like serine protease.

BACKGROUND OF THE INVENTION

Mannose-binding lectin-associate serine protease (MASP) is a member of the serine protease superfamily. Endo *et al.*, *J. Immunol.* 161, 4924-30, 1998. In humans, MASP plays a role in the host defense against pathogens via the lectin pathway, a system of complement activation. *Id.* This pathway is initiated by the binding of MASP to carbohydrates. Takahashi *et al.*, *Int. Immunol.* 11, 859-63, 1999. Because of the importance of defending the human host against a variety of pathogens, there is a need in the art to identify additional members of this protein family which can be regulated to provide therapeutic effects.

SUMMARY OF THE INVENTION

It is an object of the invention to provide reagents and methods of regulating extracellular matrix degradation. These and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention is a MASP-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2; and

the amino acid sequence shown in SEQ ID NO: 2.

- 2 -

Yet another embodiment of the invention is a method of screening for agents which decrease extracellular matrix degradation. A test compound is contacted with a MASP-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

5

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2; and

the amino acid sequence shown in SEQ ID NO: 2.

10

Binding between the test compound and the MASP-like serine protease polypeptide is detected. A test compound which binds to the MASP-like serine protease polypeptide is thereby identified as a potential agent for decreasing extracellular matrix degradation. The agent can work by decreasing the activity of the MASP-like serine protease.

15

Another embodiment of the invention is a method of screening for agents which decrease extracellular matrix degradation. A test compound is contacted with a polynucleotide encoding a MASP-like serine protease polypeptide, wherein the polynucleotide comprises a nucleotide sequence selected from the group consisting of:

20

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 1; and

25

the nucleotide sequence shown in SEQ ID NO: 1.

Binding of the test compound to the polynucleotide is detected. A test compound which binds to the polynucleotide is identified as a potential agent for decreasing extracellular matrix degradation. The agent can work by decreasing the amount of the

30

MASP-like serine protease through interacting with the MASP-like serine protease mRNA.

5 Another embodiment of the invention is a method of screening for agents which regulate extracellular matrix degradation. A test compound is contacted with a MASP-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

10 amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2; and

the amino acid sequence shown in SEQ ID NO: 2.

15 A MASP-like serine protease activity of the polypeptide is detected. A test compound which increases MASP-like serine protease activity of the polypeptide relative to MASP-like serine protease activity in the absence of the test compound is thereby identified as a potential agent for increasing extracellular matrix degradation. A test compound which decreases MASP-like serine protease activity of the polypeptide relative to MASP-like serine protease activity in the absence of the test
20 compound is thereby identified as a potential agent for decreasing extracellular matrix degradation.

25 Even another embodiment of the invention is a method of screening for agents which decrease extracellular matrix degradation. A test compound is contacted with a MASP-like serine protease product of a polynucleotide which comprises a nucleotide sequence selected from the group consisting of:

30 nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 1; and

the nucleotide sequence shown in SEQ ID NO: 1.

Binding of the test compound to the MASP-like serine protease product is detected. A test compound which binds to the MASP-like serine protease product is thereby identified as a potential agent for decreasing extracellular matrix degradation.

5

Still another embodiment of the invention is a method of reducing extracellular matrix degradation. A cell is contacted with a reagent which specifically binds to a polynucleotide encoding a MASP-like serine protease polypeptide or the product encoded by the polynucleotide, wherein the polynucleotide comprises a nucleotide sequence selected from the group consisting of:

10

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 1; and

15

the nucleotide sequence shown in SEQ ID NO: 1.

MASP-like serine protease activity in the cell is thereby decreased.

20

The invention thus provides reagents and methods for regulating MASP-like serine protease activity which can be used *inter alia*, to treat pathogenic infections.

BRIEF DESCRIPTION OF THE DRAWINGS

25

Fig. 1 shows the DNA-sequence encoding a MASP-like serine protease polypeptide (SEQ ID NO:1).

Fig. 2 shows the amino acid sequence of the of the DNA sequence of Fig. 1 (SEQ ID NO:2).

Fig. 3 shows the amino acid sequence of a protein identified by EMBL Accession No. AB009074 (SEQ ID NO:3).

30

Fig. 4 shows the amino acid sequence of pfam/hmm/trypsin (SEQ ID NO:4).

Fig. 5 shows the BLASTP alignment of SEQ ID NO:2 with SEQ ID NO: 3

Fig. 6 shows the Prosite search results

Fig. 7 shows the BLOCKS search results

Fig. 8 shows the HMMPFAM alignment of 142_2_TR3 (SEQ ID NO:2) against pfam/hmm/trypsin (SEQ ID NO:4)

5

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to an isolated polynucleotide encoding a MASP-like serine protease polypeptide and being selected from the group consisting of:

- 10 a) a polynucleotide encoding a MASP-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to

- 15 the amino acid sequence shown in SEQ ID NO: 2; and

the amino acid sequence shown in SEQ ID NO: 2;

- b) a polynucleotide comprising the sequence of SEQ ID NO: 1,
c) a polynucleotide which hybridizes under stringent conditions to a
20 polynucleotide specified in (a) and (b);
d) a polynucleotide the sequence of which deviates from the polynucleotide sequences specified in (a) to (c) due to the degeneration of the genetic code; and
e) a polynucleotide which represents a fragment, derivative or allelic variation of
25 a polynucleotide sequence specified in (a) to (d).

Furthermore, it has been discovered by the present applicant that regulators of a MASP-like serine protease, particularly a human MASP-like serine protease, can be used to regulate degradation of the extracellular matrix. Human MASP-like serine
30 protease as shown in SEQ ID NO:2 is 64% identical over 282 amino acids to the protein identified by EMBL Accession No. AB009074 (SEQ ID NO:3) and

annotated as *Triakis scyllium* mannose-binding lectin-associated serine protease (FIG. 1). Human MASP-like serine protease contains domains typical of a trypsin family serine protease (FIGS. 2-4).

5 A coding sequence for SEQ ID NO:2 is shown in SEQ ID NO:1. This coding sequence is found within genomic clones identified with GenBank Accession Nos. AC069069, AC007920, AC034190, and AC046154. An EST identified with EMBL Accession No. AW855182 (423bp) is contained within SEQ ID NO:1, indicating that human MASP-like serine protease is expressed.

10

Human MASP-like serine protease is expected to be useful for the same purposes as previously identified mannose-binding lectin-associated serine proteases (Endo *et al.*, *J. Immunol.* 161, 4924-30, 1998; Takahashi *et al.*, *Int. Immunol.* 11, 859-63, 1999; Matsushita *et al.*, *J. Immunol.* 164, 2281-84, 2000). Human MASP-like serine
15 protease is expected to be especially useful for treating diseases caused by pathogens, including but not limited to, viruses, bacteria, mycoplasma, fungi, protozoa, helminths, rickettsia, chloamydiae, parasites, prions, and the like.

Polypeptides

20

MASP-like serine protease polypeptides according to the invention comprise an amino acid sequence as shown in SEQ ID NO:2, a portion of SEQ ID NO:2 comprising at least 6, 15, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 275 contiguous amino acids, or a biologically active variant of the amino acid sequence
25 shown in SEQ ID NO:2, as defined below. A MASP-like serine protease polypeptide of the invention therefore can be a portion of a MASP-like serine protease molecule, a full-length MASP-like serine protease molecule, or a fusion protein comprising all or a portion of a MASP-like serine protease molecule.

Biologically Active Variants

MASP-like serine protease variants which are biologically active, *i.e.*, retain a MASP-like serine protease activity, also are MASP-like serine protease polypeptides. Preferably, naturally or non-naturally occurring MASP-like serine protease variants have amino acid sequences which are at least about 50, 55, 60, 65, 70, preferably about 75, 90, 96, or 98% identical to an amino acid sequence shown in SEQ ID NO:2. Percent identity between a putative MASP-like serine protease variant and an amino acid sequence of SEQ ID NO:2 is determined using the Blast2 alignment program.

Variations in percent identity can be due, for example, to amino acid substitutions, insertions, or deletions. Amino acid substitutions are defined as one for one amino acid replacements. They are conservative in nature when the substituted amino acid has similar structural and/or chemical properties. Examples of conservative replacements are substitution of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine.

Amino acid insertions or deletions are changes to or within an amino acid sequence. They typically fall in the range of about 1 to 5 amino acids. Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, such as DNASTAR software. Whether an amino acid change results in a biologically active MASP-like serine protease polypeptide can readily be determined by assaying for MASP-like serine protease activity, as is known in the art and described, for example, in Matsushita *et al.*, 2000.

Fusion Proteins

Fusion proteins are useful for generating antibodies against MASP-like serine protease amino acid sequences and for use in various assay systems. For example,

fusion proteins can be used to identify proteins which interact with portions of a MASP-like serine protease polypeptide, including its active site and fibronectin domains. Methods such as protein affinity chromatography or library-based assays for protein-protein interactions, such as the yeast two-hybrid or phage display systems, can be used for this purpose. Such methods are well known in the art and also can be used as drug screens.

A MASP-like serine protease fusion protein comprises two protein segments fused together by means of a peptide bond. Contiguous amino acids for use in a fusion protein can be selected from the amino acid sequence shown in SEQ ID NO:2 or from a biologically active variants of those sequences, such as those described above. For example, the first protein segment can comprise at least 6, 15, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 275 or more contiguous amino acids of SEQ ID NO:2 or a biologically active variant. Preferably, a fusion protein comprises the active site of the protease and/or one or both of the fibronectin domains. The first protein segment also can comprise full-length MASP-like serine protease.

The second protein segment can be a full-length protein or a protein fragment or polypeptide. Proteins commonly used in fusion protein construction include β -galactosidase, β -glucuronidase, green fluorescent protein (GFP), autofluorescent proteins, including blue fluorescent protein (BFP), glutathione-S-transferase (GST), luciferase, horseradish peroxidase (HRP), and chloramphenicol acetyltransferase (CAT). Additionally, epitope tags are used in fusion protein constructions, including histidine (His) tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Other fusion constructions can include maltose binding protein (MBP), S-tag, Lex a DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. A fusion protein also can be engineered to contain a cleavage site located between the MASP-like serine protease polypeptide-encoding sequence and the heterologous protein sequence, so that the MASP-like serine protease polypeptide can be cleaved and purified away from the heterologous moiety.

A fusion protein can be synthesized chemically, as is known in the art. Preferably, a fusion protein is produced by covalently linking two protein segments or by standard procedures in the art of molecular biology. Recombinant DNA methods can be used to prepare fusion proteins, for example, by making a DNA construct which comprises MASP-like serine protease coding sequences disclosed herein in proper reading frame with nucleotides encoding the second protein segment and expressing the DNA construct in a host cell, as is known in the art. Many kits for constructing fusion proteins are available from companies such as Promega Corporation (Madison, WI), Stratagene (La Jolla, CA), CLONTECH (Mountain View, CA), Santa Cruz Biotechnology (Santa Cruz, CA), MBL International Corporation (MIC; Watertown, MA), and Quantum Biotechnologies (Montreal, Canada; 1-888-DNA-KITS).

Identification of Species Homologs

Species homologs of human MASP-like serine protease can be obtained using MASP-like serine protease polynucleotides (described below) to make suitable probes or primers to screening cDNA expression libraries from other species, such as mice, monkeys, or yeast, identifying cDNAs which encode homologs of MASP-like serine protease, and expressing the cDNAs as is known in the art.

Polynucleotides

A MASP-like serine protease polynucleotide can be single- or double-stranded and comprises a coding sequence or the complement of a coding sequence for a MASP-like serine protease polypeptide. A partial coding sequence of a MASP-like serine protease polynucleotide is shown in SEQ ID NO:1.

Degenerate nucleotide sequences encoding human MASP-like serine protease polypeptides, as well as homologous nucleotide sequences which are at least about

50, 55, 60, 65, 70, preferably about 75, 90, 96, or 98% identical to the MASP-like serine protease coding sequences nucleotide sequence shown in SEQ ID NO:1 also are MASP-like serine protease polynucleotides. Percent sequence identity between the sequences of two polynucleotides is determined using computer programs such as
5 ALIGN which employ the FASTA algorithm, using an affine gap search with a gap open penalty of -12 and a gap extension penalty of -2. Complementary DNA (cDNA) molecules, species homologs, and variants of MASP-like serine protease polynucleotides which encode biologically active MASP-like serine protease polypeptides also are MASP-like serine protease polynucleotides.

Identification of Variants and Homologs

Variants and homologs of the MASP-like serine protease polynucleotides disclosed above also are MASP-like serine protease polynucleotides. Typically, homologous
15 MASP-like serine protease polynucleotide sequences can be identified by hybridization of candidate polynucleotides to known MASP-like serine protease polynucleotides under stringent conditions, as is known in the art. For example, using the following wash conditions--2X SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7.0), 0.1% SDS, room temperature twice, 30 minutes each; then 2X SSC, 0.1%
20 SDS, 50 °C once, 30 minutes; then 2X SSC, room temperature twice, 10 minutes each--homologous sequences can be identified which contain at most about 25-30% basepair mismatches. More preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches.

25 Species homologs of the MASP-like serine protease polynucleotides disclosed herein can be identified by making suitable probes or primers and screening cDNA expression libraries from other species, such as mice, monkeys, or yeast. Human variants of MASP-like serine protease polynucleotides can be identified, for example, by screening human cDNA expression libraries. It is well known that the
30 T_m of a double-stranded DNA decreases by 1-1.5 °C with every 1% decrease in homology (Bonner *et al.*, *J. Mol. Biol.* 81, 123 (1973). Variants of human MASP-

like serine protease polynucleotides or MASP-like serine protease polynucleotides of other species can therefore be identified, for example, by hybridizing a putative homologous MASP-like serine protease polynucleotide with a polynucleotide having a nucleotide sequence of SEQ ID NO:1 or an ephrin-like serine protease coding sequence of SEQ ID NO: 3 to form a test hybrid. The melting temperature of the test hybrid is compared with the melting temperature of a hybrid comprising MASP-like serine protease polynucleotides having perfectly complementary nucleotide sequences, and the number or percent of basepair mismatches within the test hybrid is calculated.

Nucleotide sequences which hybridize to MASP-like serine protease polynucleotides or their complements following stringent hybridization and/or wash conditions are also MASP-like serine protease polynucleotides. Stringent wash conditions are well known and understood in the art and are disclosed, for example, in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed., 1989, at pages 9.50-9.51.

Typically, for stringent hybridization conditions a combination of temperature and salt concentration should be chosen that is approximately 12-20 °C below the calculated T_m of the hybrid under study. The T_m of a hybrid between a MASP-like serine protease polynucleotide having a coding sequence disclosed herein and a polynucleotide sequence which is at least about 50, preferably about 75, 90, 96, or 98% identical to that nucleotide sequence can be calculated, for example, using the equation of Bolton and McCarthy, *Proc. Natl. Acad. Sci. U.S.A.* 48, 1390 (1962):

$$T_m = 81.5\text{ }^{\circ}\text{C} - 16.6(\log_{10} [\text{Na}^+]) + 0.41(\%G + C) - 0.63(\%\text{formamide}) - 600/l,$$

where l = the length of the hybrid in basepairs.

Stringent wash conditions include, for example, 4X SSC at 65 °C, or 50% formamide, 4X SSC at 42 °C, or 0.5X SSC, 0.1% SDS at 65 °C. Highly stringent wash conditions include, for example, 0.2X SSC at 65 °C.

Preparation of Polynucleotides

- A naturally occurring MASP-like serine protease polynucleotide can be isolated free of other cellular components such as membrane components, proteins, and lipids.
- 5 Polynucleotides can be made by a cell and isolated using standard nucleic acid purification techniques, synthesized using an amplification technique, such as the polymerase chain reaction (PCR), or synthesized using an automatic synthesizer. Methods for isolating polynucleotides are routine and are known in the art. Any such technique for obtaining a polynucleotide can be used to obtain isolated MASP-like
- 10 serine protease polynucleotides. For example, restriction enzymes and probes can be used to isolate polynucleotide fragments which comprise MASP-like serine protease nucleotide sequences. Isolated polynucleotides are in preparations which are free or at least 70, 80, or 90% free of other molecules.
- 15 MASP-like serine protease cDNA molecules can be made with standard molecular biology techniques, using MASP-like serine protease mRNA as a template. MASP-like serine protease cDNA molecules can thereafter be replicated using molecular biology techniques known in the art and disclosed in manuals such as Sambrook *et al.* (1989). An amplification technique, such as PCR, can be used to obtain
- 20 additional copies of MASP-like serine protease polynucleotides, using either human genomic DNA or cDNA as a template.
- Alternatively, synthetic chemistry techniques can be used to synthesize MASP-like serine protease polynucleotides. The degeneracy of the genetic code allows alternate
- 25 nucleotide sequences to be synthesized which will encode a MASP-like serine protease polypeptide having, for example, the amino acid sequence shown in SEQ ID NO:2 or a biologically active variant of that sequence.

Obtaining Full-Length Polynucleotides

5 The partial sequence of SEQ ID NO:1 or its complement can be used to identify the corresponding full length gene from which they were derived. The partial sequences can be nick-translated or end-labeled with ^{32}P using polynucleotide kinase using labeling methods known to those with skill in the art (BASIC METHODS IN MOLECULAR BIOLOGY, Davis *et al.*, eds., Elsevier Press, N.Y., 1986). A lambda library prepared from human tissue can be directly screened with the labeled sequences of interest or the library can be converted en masse to pBluescript (Stratagene Cloning Systems, La Jolla, Calif. 92037) to facilitate bacterial colony screening (see Sambrook *et al.*, 1989, pg. 1.20).

15 Both methods are well known in the art. Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured, and the DNA is fixed to the filters. The filters are hybridized with the labeled probe using hybridization conditions described by Davis *et al.*, 1986. The partial sequences, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive colony or plaque. The colonies or plaques are selected and expanded, and the DNA is isolated from the colonies for further analysis and sequencing.

25 Positive cDNA clones are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the partial sequence and the other primer from the vector. Clones with a larger vector-insert PCR product than the original partial sequence are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size determined from Northern blot Analysis.

30

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined, for example after exonuclease III digestion (McCombie *et al.*, *Methods* 3, 33-40, 1991). A series of deletion clones are generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

Various PCR-based methods can be used to extend the nucleic acid sequences encoding the disclosed portions of human MASP-like serine protease to detect upstream sequences such as promoters and regulatory elements. For example, restriction-site PCR uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, *PCR Methods Applic.* 2, 318-322, 1993). Genomic DNA is first amplified in the presence of a primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR also can be used to amplify or extend sequences using divergent primers based on a known region (Triglia *et al.*, *Nucleic Acids Res.* 16, 8186, 1988). Primers can be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, Minn.), to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68 - 72 °C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which can be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom *et al.*, *PCR Methods Applic.* 1, 111-119,

1991). In this method, multiple restriction enzyme digestions and ligations are used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

5 Another method which can be used to retrieve unknown sequences is that of Parker *et al.*, *Nucleic Acids Res.* 19, 3055-3060, 1991. Additionally, PCR, nested primers, and PROMOTERFINDER libraries (CLONTECH, Palo Alto, Calif.) can be used to walk genomic DNA. This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

10

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an
15 oligo d(T) library does not yield a full-length cDNA. Genomic libraries can be useful for extension of sequence into 5' non-transcribed regulatory regions.

20

Commercially available capillary electrophoresis systems can be used to analyze the size or confirm the nucleotide sequence of PCR or sequencing products. For example, capillary sequencing can employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity can be converted to electrical signal using appropriate software (*e.g.* GENOTYPER and Sequence NAVIGATOR, Perkin Elmer), and the
25 entire process from loading of samples to computer analysis and electronic data display can be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

Obtaining Polypeptides

MASP-like serine protease polypeptides can be obtained, for example, by purification from human cells, by expression of MASP-like serine protease polynucleotides, or by direct chemical synthesis.

Protein Purification

MASP-like serine protease polypeptides can be purified from human cells, such as primary tumor cells, metastatic cells, or cancer cell lines (e.g., colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, or the H392 glioblastoma cell line). Carcinoma of the lung is an especially useful source of MASP-like serine protease polypeptides. A purified MASP-like serine protease polypeptide is separated from other compounds which normally associate with the MASP-like serine protease polypeptide in the cell, such as certain proteins, carbohydrates, or lipids, using methods well-known in the art. Such methods include, but are not limited to, size exclusion chromatography, ammonium sulfate fractionation, ion exchange chromatography, affinity chromatography, and preparative gel electrophoresis. A preparation of purified MASP-like serine protease polypeptides is at least 80% pure; preferably, the preparations are 90%, 95%, or 99% pure. Purity of the preparations can be assessed by any means known in the art, such as SDS-polyacrylamide gel electrophoresis. Enzymatic activity of the purified preparations can be assayed, for example, as described in Matsushita *et al.*, 2000.

Expression of Polynucleotides

To express a MASP-like serine protease polypeptide, a MASP-like serine protease polynucleotide can be inserted into an expression vector which contains the necessary elements for the transcription and translation of the inserted coding

sequence. Methods which are well known to those skilled in the art can be used to construct expression vectors containing sequences encoding MASP-like serine protease polypeptides and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic
5 techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook *et al.* (1989) and Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1989.

A variety of expression vector/host systems can be utilized to contain and express
10 sequences encoding a MASP-like serine protease polypeptide. These include, but are not limited to, microorganisms, such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors, insect cell systems infected with virus expression vectors (*e.g.*, baculovirus), plant cell systems transformed with virus expression vectors (*e.g.*,
15 cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids), or animal cell systems.

The control elements or regulatory sequences are those non-translated regions of the vector -- enhancers, promoters, 5' and 3' untranslated regions -- which interact with
20 host cellular proteins to carry out transcription and translation. Such elements can vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, can be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the
25 BLUESCRIPT phagemid (Stratagene, LaJolla, Calif.) or pSPORT1 plasmid (Life Technologies) and the like can be used. The baculovirus polyhedrin promoter can be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (*e.g.*, heat shock, RUBISCO, and storage protein genes) or from plant viruses (*e.g.*, viral promoters or leader sequences) can be cloned into the vector. In
30 mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple

copies of a nucleotide sequence encoding a MASP-like serine protease polypeptide, vectors based on SV40 or EBV can be used with an appropriate selectable marker.

Bacterial and Yeast Expression Systems

5

In bacterial systems, a number of expression vectors can be selected depending upon the use intended for the MASP-like serine protease polypeptide. For example, when a large quantity of a MASP-like serine protease polypeptide is needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified can be used. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the MASP-like serine protease polypeptide can be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced. pIN vectors (Van Heeke & Schuster, *J. Biol. Chem.* 264, 5503-5509, 1989 or pGEX vectors (Promega, Madison, Wis.) can be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems can be designed to include heparin, thrombin, or Factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

10
15
20

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH can be used. For reviews, see Ausubel *et al.* (1989) and Grant *et al.*, *Methods Enzymol.* 153, 516-544, 1987.

25

Plant and Insect Expression Systems

If plant expression vectors are used, the expression of sequences encoding MASP-like serine protease polypeptides can be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV can be used alone or in combination with the omega leader sequence from TMV (Takamatsu *EMBO J.* 6, 307-311, 1987). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters can be used (Coruzzi *et al.*, *EMBO J.* 3, 1671-1680, 1984; Broglie *et al.*, *Science* 224, 838-843, 1984; Winter *et al.*, *Results Probl. Cell Differ.* 17, 85-105, 1991). These constructs can be introduced into plant cells by direct DNA transformation or by pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs or Murray, in MCGRAW HILL YEARBOOK OF SCIENCE AND TECHNOLOGY, McGraw Hill, New York, N.Y., pp. 191-196, 1992).

An insect system also can be used to express a MASP-like serine protease polypeptide. For example, in one such system *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. Sequences encoding MASP-like serine protease polypeptides can be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of MASP-like serine protease polypeptides will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses can then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which MASP-like serine protease polypeptides can be expressed (Engelhard *et al.*, *Proc. Nat. Acad. Sci.* 91, 3224-3227, 1994).

Mammalian Expression Systems

A number of viral-based expression systems can be utilized in mammalian host cells. For example, if an adenovirus is used as an expression vector, sequences encoding
5 MASP-like serine protease polypeptides can be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome can be used to obtain a viable virus which is capable of expressing a MASP-like serine
10 protease polypeptide in infected host cells (Logan & Shenk, *Proc. Natl. Acad. Sci.* 81, 3655-3659, 1984). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, can be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) also can be used to deliver larger fragments
15 of DNA than can be contained and expressed in a plasmid. HACs of 6M to 10M are constructed and delivered to cells via conventional delivery methods (e.g., liposomes, polycationic amino polymers, or vesicles).

Specific initiation signals also can be used to achieve more efficient translation of
20 sequences encoding MASP-like serine protease polypeptides. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding a MASP-like serine protease polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases
25 where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals (including the ATG initiation codon) should be provided. The initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be
30 enhanced by the inclusion of enhancers which are appropriate for the particular cell

system which is used (see Scharf *et al.*, *Results Probl. Cell Differ.* 20, 125-162, 1994).

Host Cells

5

A host cell strain can be chosen for its ability to modulate the expression of the inserted sequences or to process an expressed MASP-like serine protease polypeptide in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the polypeptide also can be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (*e.g.*, CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; 10801 University Boulevard, Manassas, VA 20110-2209) and can be chosen to ensure the correct modification and processing of the foreign protein.

Stable expression is preferred for long-term, high-yield production of recombinant proteins. For example, cell lines which stably express MASP-like serine protease polypeptides can be transformed using expression vectors which can contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells can be allowed to grow for 1-2 days in an enriched medium before they are switched to a selective medium. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced MASP-like serine protease sequences. Resistant clones of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems can be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase

(Wigler *et al.*, *Cell* 11, 223-32, 1977) and adenine phosphoribosyltransferase (Lowy *et al.*, *Cell* 22, 817-23, 1980). Genes which can be employed in *tk* or *aprt* cells, respectively. Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate (Wigler *et al.*, *Proc. Natl. Acad. Sci.* 77, 3567-70, 1980); *npt* confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al.*, *J. Mol. Biol.* 150, 1-14, 1981); and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murray, 1992 *supra*). Additional selectable genes have been described, for example *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci.* 85, 8047-51, 1988). Visible markers such as anthocyanins, β -glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, can be used to identify transformants and to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes *et al.*, *Methods Mol. Biol.* 55, 121-131, 1995).

Detecting Expression of Polypeptides

Although the presence of marker gene expression suggests that the MASP-like serine protease polynucleotide is also present, its presence and expression may need to be confirmed. For example, if a sequence encoding a MASP-like serine protease polypeptide is inserted within a marker gene sequence, transformed cells containing sequences which encode a MASP-like serine protease polypeptide can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a MASP-like serine protease polypeptide under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the MASP-like serine protease polynucleotide.

Alternatively, host cells which contain a MASP-like serine protease polynucleotide and which express a MASP-like serine protease polypeptide can be identified by a

variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip-based technologies for the detection and/or quantification of nucleic acid or protein.

5

The presence of a polynucleotide sequence encoding a MASP-like serine protease polypeptide can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding a MASP-like serine protease polypeptide. Nucleic acid amplification-based assays involve the use of oligonucleotides selected from sequences encoding a MASP-like serine protease polypeptide to detect transformants which contain a MASP-like serine protease polynucleotide.

10

A variety of protocols for detecting and measuring the expression of a MASP-like serine protease polypeptide, using either polyclonal or monoclonal antibodies specific for the polypeptide, are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay using monoclonal antibodies reactive to two non-interfering epitopes on a MASP-like serine protease polypeptide can be used, or a competitive binding assay can be employed. These and other assays are described in Hampton *et al.*, SEROLOGICAL METHODS: A LABORATORY MANUAL, APS Press, St. Paul, Minn., 1990) and Maddox *et al.*, *J. Exp. Med.* 158, 1211-1216, 1983).

15

20

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding MASP-like serine protease polypeptides include oligo-labeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, sequences encoding a MASP-like serine protease polypeptide can be cloned into a vector for the production of an mRNA probe. Such vectors are

25

30

known in the art, are commercially available, and can be used to synthesize RNA probes *in vitro* by addition of labeled nucleotides and an appropriate RNA polymerase, such as T7, T3, or SP6. These procedures can be conducted using a variety of commercially available kits (Amersham Pharmacia Biotech, Promega, and US Biochemical). Suitable reporter molecules or labels which can be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

10 Expression and Purification of Polypeptides

Host cells transformed with nucleotide sequences encoding a MASP-like serine protease polypeptide can be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The polypeptide produced by a transformed cell can be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode MASP-like serine protease polypeptides can be designed to contain signal sequences which direct secretion of MASP-like serine protease polypeptides through a prokaryotic or eukaryotic cell membrane.

20

Other constructions can be used to join a sequence encoding a MASP-like serine protease polypeptide to a nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor Xa or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the MASP-like serine protease polypeptide can be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a

30

5 MASP-like serine protease polypeptide and 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography as described in Porath *et al.*, *Prot. Exp. Purif.* 3, 263-281, 1992), while the enterokinase cleavage site provides a means for purifying the MASP-like serine protease polypeptide from the fusion protein. Vectors which contain fusion proteins are disclosed in Kroll *et al.*, *DNA Cell Biol.* 12, 441-453, 1993).

Chemical Synthesis

10

Sequences encoding a MASP-like serine protease polypeptide can be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers *et al.*, *Nucl. Acids Res. Symp. Ser.* 215-223, 1980; Horn *et al.* *Nucl. Acids Res. Symp. Ser.* 225-232, 1980). Alternatively, a MASP-like serine protease polypeptide itself can be produced using chemical methods to synthesize its amino acid sequence. For example, MASP-like serine protease polypeptides can be produced by direct peptide synthesis using solid-phase techniques (Merrifield, *J. Am. Chem. Soc.* 85, 2149-2154, 1963; Roberge *et al.*, *Science* 269, 202-204, 1995). Protein synthesis can be performed using manual techniques or by automation. Automated synthesis can be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of MASP-like serine protease polypeptides can be separately synthesized and combined using chemical methods to produce a full-length molecule.

20

25 The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, *PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES*, WH Freeman and Co., New York, N.Y., 1983). The composition of a synthetic MASP-like serine protease polypeptide can be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure; *see* Creighton, *supra*). Additionally, any portion of the amino acid sequence of the MASP-like serine protease polypeptide can be altered during direct synthesis and/or

30

combined using chemical methods with sequences from other proteins to produce a variant polypeptide or a fusion protein.

Production of Altered Polypeptides

5

As will be understood by those of skill in the art, it may be advantageous to produce MASP-like serine protease polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

10

The nucleotide sequences disclosed herein can be engineered using methods generally known in the art to alter MASP-like serine protease polypeptide-encoding sequences for a variety of reasons, including modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides can be used to engineer the nucleotide sequences. For example, site-directed mutagenesis can be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

15

20

Antibodies

25

30

Any type of antibody known in the art can be generated to bind specifically to an epitope of a MASP-like serine protease polypeptide. "Antibody" as used herein includes intact immunoglobulin molecules, as well as fragments thereof, such as Fab, F(ab')₂, and Fv, which are capable of binding an epitope of a MASP-like serine protease polypeptide. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. However, epitopes which involve non-contiguous amino acids may require more, e.g., at least 15, 25, or 50 amino acids.

An antibody which specifically binds to an epitope of a MASP-like serine protease polypeptide can be used therapeutically, as well as in immunochemical assays, including but not limited to Western blots, ELISAs, radioimmunoassays, immuno-
5 histochemical assays, immunoprecipitations, or other immunochemical assays known in the art. Various immunoassays can be used to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays are well known in the art. Such immunoassays typically involve the measurement of complex formation between an immunogen and an antibody which
10 specifically binds to the immunogen.

Typically, an antibody which specifically binds to a MASP-like serine protease polypeptide provides a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in an immunochemical
15 assay. Preferably, antibodies which specifically bind to MASP-like serine protease polypeptides do not detect other proteins in immunochemical assays and can immunoprecipitate a MASP-like serine protease polypeptide from solution.

MASP-like serine protease polypeptides can be used to immunize a mammal, such as
20 a mouse, rat, rabbit, guinea pig, monkey, or human, to produce polyclonal antibodies. If desired, a MASP-like serine protease polypeptide can be conjugated to a carrier protein, such as bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin. Depending on the host species, various adjuvants can be used to increase the immunological response. Such adjuvants include, but are not limited to,
25 Freund's adjuvant, mineral gels (*e.g.*, aluminum hydroxide), and surface active substances (*e.g.* lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol). Among adjuvants used in humans, BCG (*bacilli Calmette-Guerin*) and *Corynebacterium parvum* are especially useful.

30 Monoclonal antibodies which specifically bind to a MASP-like serine protease polypeptide can be prepared using any technique which provides for the production

of antibody molecules by continuous cell lines in culture. These techniques include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler *et al.*, *Nature* 256, 495-497, 1985; Kozbor *et al.*, *J. Immunol. Methods* 81, 31-42, 1985; Cote *et al.*, *Proc. Natl. Acad. Sci.* 80, 2026-2030, 1983; Cole *et al.*, *Mol. Cell Biol.* 62, 109-120, 1984).

In addition, techniques developed for the production of "chimeric antibodies," the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison *et al.*, *Proc. Natl. Acad. Sci.* 81, 6851-6855, 1984; Neuberger *et al.*, *Nature* 312, 604-608, 1984; Takeda *et al.*, *Nature* 314, 452-454, 1985). Monoclonal and other antibodies also can be "humanized" to prevent a patient from mounting an immune response against the antibody when it is used therapeutically. Such antibodies may be sufficiently similar in sequence to human antibodies to be used directly in therapy or may require alteration of a few key residues. Sequence differences between rodent antibodies and human sequences can be minimized by replacing residues which differ from those in the human sequences by site directed mutagenesis of individual residues or by grating of entire complementarity determining regions. Alternatively, one can produce humanized antibodies using recombinant methods, as described in GB2188638B. Antibodies which specifically bind to a MASP-like serine protease polypeptide can contain antigen binding sites which are either partially or fully humanized, as disclosed in U.S. 5,565,332.

Alternatively, techniques described for the production of single chain antibodies can be adapted using methods known in the art to produce single chain antibodies which specifically bind to MASP-like serine protease polypeptides. Antibodies with related specificity, but of distinct idiotypic composition, can be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, *Proc. Natl. Acad. Sci.* 88, 11120-23, 1991).

Single-chain antibodies also can be constructed using a DNA amplification method, such as PCR, using hybridoma cDNA as a template (Thirion *et al.*, 1996, *Eur. J. Cancer Prev.* 5, 507-11). Single-chain antibodies can be mono- or bispecific, and can be bivalent or tetravalent. Construction of tetravalent, bispecific single-chain antibodies is taught, for example, in Coloma & Morrison, 1997, *Nat. Biotechnol.* 15, 159-63. Construction of bivalent, bispecific single-chain antibodies is taught in Mallender & Voss, 1994, *J. Biol. Chem.* 269, 199-206.

A nucleotide sequence encoding a single-chain antibody can be constructed using manual or automated nucleotide synthesis, cloned into an expression construct using standard recombinant DNA methods, and introduced into a cell to express the coding sequence, as described below. Alternatively, single-chain antibodies can be produced directly using, for example, filamentous phage technology. Verhaar *et al.*, 1995, *Int. J. Cancer* 61, 497-501; Nicholls *et al.*, 1993, *J. Immunol. Meth.* 165, 81-91.

Antibodies which specifically bind to MASP-like serine protease polypeptides also can be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi *et al.*, *Proc. Natl. Acad. Sci.* 86, 3833-3837, 1989; Winter *et al.*, *Nature* 349, 293-299, 1991).

Other types of antibodies can be constructed and used therapeutically in methods of the invention. For example, chimeric antibodies can be constructed as disclosed in WO 93/03151. Binding proteins which are derived from immunoglobulins and which are multivalent and multispecific, such as the "diabodies" described in WO 94/13804, also can be prepared.

Antibodies of the invention can be purified by methods well known in the art. For example, antibodies can be affinity purified by passage over a column to which a

MASP-like serine protease polypeptide is bound. The bound antibodies can then be eluted from the column using a buffer with a high salt concentration.

Antisense Oligonucleotides

5

Antisense oligonucleotides are nucleotide sequences which are complementary to a specific DNA or RNA sequence. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form complexes and block either transcription or translation. Preferably, an antisense oligonucleotide is at least 11 nucleotides in length, but can be at least 12, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides long. Longer sequences also can be used. Antisense oligonucleotide molecules can be provided in a DNA construct and introduced into a cell as described above to decrease the level of MASP-like serine protease gene products in the cell.

15

Antisense oligonucleotides can be deoxyribonucleotides, ribonucleotides, or a combination of both. Oligonucleotides can be synthesized manually or by an automated synthesizer, by covalently linking the 5' end of one nucleotide with the 3' end of another nucleotide with non-phosphodiester internucleotide linkages such as alkylphosphonates, phosphorothioates, phosphorodithioates, alkylphosphonothioates, alkylphosphonates, phosphoramidates, phosphate esters, carbamates, acetamidate, carboxymethyl esters, carbonates, and phosphate triesters. See Brown, *Meth. Mol. Biol.* 20, 1-8, 1994; Sonveaux, *Meth. Mol. Biol.* 26, 1-72, 1994; Uhlmann *et al.*, *Chem. Rev.* 90, 543-583, 1990.

25

Modifications of MASP-like serine protease gene expression can be obtained by designing antisense oligonucleotides which will form duplexes to the control, 5', or regulatory regions of the MASP-like serine protease gene. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition

30

of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or chaperons. Therapeutic advances using triplex DNA have been described in the literature (*e.g.*, Gee *et al.*, in Huber & Carr, MOLECULAR AND IMMUNOLOGIC APPROACHES, Futura Publishing Co., Mt. Kisco, N.Y., 1994). An
5 antisense oligonucleotide also can be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Precise complementarity is not required for successful duplex formation between an antisense oligonucleotide and the complementary sequence of a MASP-like serine
10 protease polynucleotide. Antisense oligonucleotides which comprise, for example, 2, 3, 4, or 5 or more stretches of contiguous nucleotides which are precisely complementary to a MASP-like serine protease polynucleotide, each separated by a stretch of contiguous nucleotides which are not complementary to adjacent MASP-like serine protease nucleotides, can provide targeting specificity for MASP-like
15 serine protease mRNA. Preferably, each stretch of complementary contiguous nucleotides is at least 4, 5, 6, 7, or 8 or more nucleotides in length. Non-complementary intervening sequences are preferably 1, 2, 3, or 4 nucleotides in length. One skilled in the art can easily use the calculated melting point of an antisense-sense pair to determine the degree of mismatching which will be tolerated between a particular
20 antisense oligonucleotide and a particular MASP-like serine protease polynucleotide sequence.

Antisense oligonucleotides can be modified without affecting their ability to hybridize to a MASP-like serine protease polynucleotide. These modifications can
25 be internal or at one or both ends of the antisense molecule. For example, internucleoside phosphate linkages can be modified by adding cholesteryl or diamine moieties with varying numbers of carbon residues between the amino groups and terminal ribose. Modified bases and/or sugars, such as arabinose instead of ribose, or a 3', 5'-substituted oligonucleotide in which the 3' hydroxyl group or the 5' phosphate group are substituted, also can be employed in a modified antisense oligo-
30 nucleotide. These modified oligonucleotides can be prepared by methods well

known in the art. See, e.g., Agrawal *et al.*, *Trends Biotechnol.* 10, 152-158, 1992; Uhlmann *et al.*, *Chem. Rev.* 90, 543-584, 1990; Uhlmann *et al.*, *Tetrahedron. Lett.* 215, 3539-3542, 1987.

5 Ribozymes

Ribozymes are RNA molecules with catalytic activity. See, e.g., Cech, *Science* 236, 1532-1539; 1987; Cech, *Ann. Rev. Biochem.* 59, 543-568; 1990, Cech, *Curr. Opin. Struct. Biol.* 2, 605-609; 1992, Couture & Stinchcomb, *Trends Genet.* 12, 510-515, 10 1996. Ribozymes can be used to inhibit gene function by cleaving an RNA sequence, as is known in the art (e.g., Haseloff *et al.*, U.S. Patent 5,641,673). The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples include engineered hammerhead motif ribozyme molecules that can 15 specifically and efficiently catalyze endonucleolytic cleavage of specific nucleotide sequences.

The coding sequence of a MASP-like serine protease polynucleotide can be used to generate ribozymes which will specifically bind to mRNA transcribed from the 20 MASP-like serine protease polynucleotide. Methods of designing and constructing ribozymes which can cleave other RNA molecules in trans in a highly sequence specific manner have been developed and described in the art (see Haseloff *et al.* *Nature* 334, 585-591, 1988). For example, the cleavage activity of ribozymes can be targeted to specific RNAs by engineering a discrete "hybridization" region into the 25 ribozyme. The hybridization region contains a sequence complementary to the target RNA and thus specifically hybridizes with the target (see, for example, Gerlach *et al.*, EP 321,201).

Specific ribozyme cleavage sites within a MASP-like serine protease RNA target are 30 initially identified by scanning the RNA molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short

RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the MASP-like serine protease target RNA containing the cleavage site can be evaluated for secondary structural features which may render the target inoperable. The suitability of candidate targets also can be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays. Longer complementary sequences can be used to increase the affinity of the hybridization sequence for the target. The hybridizing and cleavage regions of the ribozyme can be integrally related; thus, upon hybridizing to the MASP-like serine protease target RNA through the complementary regions, the catalytic region of the ribozyme can cleave the target.

Ribozymes can be introduced into cells as part of a DNA construct. Mechanical methods, such as microinjection, liposome-mediated transfection, electroporation, or calcium phosphate precipitation, can be used to introduce a ribozyme-containing DNA construct into cells in which it is desired to decrease MASP-like serine protease expression. Alternatively, if it is desired that the cells stably retain the DNA construct, it can be supplied on a plasmid and maintained as a separate element or integrated into the genome of the cells, as is known in the art. The DNA construct can include transcriptional regulatory elements, such as a promoter element, an enhancer or UAS element, and a transcriptional terminator signal, for controlling transcription of ribozymes in the cells.

As taught in Haseloff *et al.*, U.S. Patent 5,641,673, ribozymes can be engineered so that ribozyme expression will occur in response to factors which induce expression of a target gene. Ribozymes also can be engineered to provide an additional level of regulation, so that destruction of MASP-like serine protease mRNA occurs only when both a ribozyme and a target gene are induced in the cells.

Identification of Target and Pathway Genes and Proteins

Described herein are methods for the identification of genes whose products interact with human matriptase. Such genes may represent genes which are differentially expressed in pathogenic infections. Further, such genes may represent genes which are differentially regulated in response to manipulations relevant to the progression or treatment of such diseases. Such differentially expressed genes may represent “target” and/or “fingerprint” genes. Methods for the identification of such differentially expressed genes are described below. Methods for the further characterization of such differentially expressed genes, and for their identification as target and/or fingerprint genes also are described below.

In addition, methods are described for the identification of genes, termed “pathway genes,” which are involved in pathogenic infections. “Pathway gene,” as used herein, refers to a gene whose gene product exhibits the ability to interact with gene products involved in these disorders. A pathway gene may be differentially expressed and, therefore, may have the characteristics of a target and/or fingerprint gene.

“Differential expression” refers to both quantitative as well as qualitative differences in a gene’s temporal and/or tissue expression pattern. Thus, a differentially expressed gene may qualitatively have its expression activated or completely inactivated in normal versus diseased states, or under control versus experimental conditions. Such a qualitatively regulated gene will exhibit an expression pattern within a given tissue or cell type which is detectable in either normal or diseased subjects, but is not detectable in both. Alternatively, such a qualitatively regulated gene will exhibit an expression pattern within a given tissue or cell type which is detectable in either control or experimental subjects, but is not detectable in both. “Detectable” refers to an RNA expression pattern which is detectable via the standard techniques of differential display, RT-PCR and/or Northern analyses, which are well known to those of skill in the art.

A differentially expressed gene may have its expression modulated, *i.e.*, quantitatively increased or decreased, in normal versus diseased states, or under control versus experimental conditions. The degree to which expression differs in normal versus body weight disorder or control versus experimental states need only be large enough to be visualized via standard characterization techniques, such as, for example, the differential display technique described below. Other such standard characterization techniques by which expression differences may be visualized include but are not limited to, quantitative RT (reverse transcriptase) PCR and Northern analyses.

Differentially expressed genes may be further described as target genes and/or fingerprint genes. "Fingerprint gene" refers to a differentially expressed gene whose expression pattern may be utilized as part of a prognostic or diagnostic evaluation, or which, alternatively, may be used in methods for identifying compounds useful for the treatment of various disorders. A fingerprint gene may also have the characteristics of a target gene or a pathway gene.

"Target gene" refers to a differentially expressed gene involved in pathogenic infections by which modulation of the level of target gene expression or of target gene product activity may act to ameliorate symptoms. A target gene may also have the characteristics of a fingerprint gene and/or a pathway gene.

Identification of Differentially Expressed Genes

A variety of methods may be utilized for the identification of genes which are involved in pathogenic infections. To identify differentially expressed genes, RNA, either total or mRNA, may be isolated from one or more tissues of the subjects utilized in paradigms such as those described above. RNA samples are obtained from tissues of experimental subjects and from corresponding tissues of control subjects. Any RNA isolation technique which does not select against the isolation of

mRNA may be utilized for the purification of such RNA samples. See, for example, Ausubel *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, Inc. New York, 1987-1993. Large numbers of tissue samples may readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski, U.S. Patent 4,843,155.

Transcripts within the collected RNA samples which represent RNA produced by differentially expressed genes may be identified by utilizing a variety of methods which are well known to those of skill in the art. For example, differential screening (Tedder *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 85, 208-12, 1988), subtractive hybridization (Hedrick *et al.*, *Nature* 308, 149-53; Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88, 2825, 1984), and, preferably, differential display (Liang & Pardee, *Science* 257, 967-71, 1992; U.S. Patent 5,262,311), may be utilized to identify nucleic acid sequences derived from genes that are differentially expressed.

Differential screening involves the duplicate screening of a cDNA library in which one copy of the library is screened with a total cell cDNA probe corresponding to the mRNA population of one cell type while a duplicate copy of the cDNA library is screened with a total cDNA probe corresponding to the mRNA population of a second cell type. For example, one cDNA probe may correspond to a total cell cDNA probe of a cell type or tissue derived from a control subject, while the second cDNA probe may correspond to a total cell cDNA probe of the same cell type or tissue derived from an experimental subject. Those clones which hybridize to one probe but not to the other potentially represent clones derived from genes differentially expressed in the cell type of interest in control versus experimental subjects.

Subtractive hybridization techniques generally involve the isolation of mRNA taken from two different sources, *e.g.*, control and experimental tissue or cell type, the hybridization of the mRNA or single-stranded cDNA reverse-transcribed from the

isolated mRNA, and the removal of all hybridized, and therefore double-stranded, sequences. The remaining non-hybridized, single-stranded cDNAs, potentially represent clones derived from genes that are differentially expressed in the two mRNA sources. Such single-stranded cDNAs are then used as the starting material
5 for the construction of a library comprising clones derived from differentially expressed genes.

The differential display technique describes a procedure, utilizing the well known polymerase chain reaction (PCR; the experimental embodiment set forth in Mullis,
10 U.S. Patent 4,683,202), which allows for the identification of sequences derived from genes which are differentially expressed. First, isolated RNA is reverse-transcribed into single-stranded cDNA, utilizing standard techniques which are well known to those of skill in the art. Primers for the reverse transcriptase reaction may include, but are not limited to, oligo dT-containing primers.

15 Next, this technique uses pairs of PCR primers, as described below, which allow for the amplification of clones representing a random subset of the RNA transcripts present within any given cell. Utilizing different pairs of primers allows each of the mRNA transcripts present in a cell to be amplified. Among such amplified transcripts
20 may be identified those which have been produced from differentially expressed genes.

The 3' oligonucleotide primer of the primer pairs may contain an oligo dT stretch of 10-13, preferably 11, dT nucleotides at its 5' end, which hybridizes to the poly(A)
25 tail of mRNA or to the complement of a cDNA reverse transcribed from an mRNA poly(A) tail. Second, in order to increase the specificity of the 3' primer, the primer may contain one or more, preferably two, additional nucleotides at its 3' end. Because, statistically, only a subset of the mRNA derived sequences present in the sample of interest will hybridize to such primers, the additional nucleotides allow the
30 primers to amplify only a subset of the mRNA derived sequences present in the sample of interest. This is preferred in that it allows more accurate and complete

visualization and characterization of each of the bands representing amplified sequences.

5 The 5' primer may contain a nucleotide sequence expected, statistically, to have the ability to hybridize to cDNA sequences derived from the tissues of interest. The nucleotide sequence may be an arbitrary one, and the length of the 5' oligonucleotide primer may range from about 9 to about 15 nucleotides, with about 13 nucleotides being preferred. Arbitrary primer sequences cause the lengths of the amplified partial cDNAs produced to be variable, thus allowing different clones to be separated by
10 using standard denaturing sequencing gel electrophoresis.

PCR reaction conditions should be chosen which optimize amplified product yield and specificity, and, additionally, produce amplified products of lengths which may be resolved utilizing standard gel electrophoresis techniques. Such reaction conditions are well known to those of skill in the art, and important reaction parameters include, for example, length and nucleotide sequence of oligonucleotide primers as
15 discussed above, and annealing and elongation step temperatures and reaction times.

The pattern of clones resulting from the reverse transcription and amplification of the mRNA of two different cell types is displayed via sequencing gel electrophoresis and compared. Differentially expressed genes are indicated by differences in the two banding patterns.
20

Once potentially differentially expressed gene sequences have been identified via bulk techniques such as, for example, those described above, the differential expression of such putatively differentially expressed genes should be corroborated. Corroboration may be accomplished via, for example, such well known techniques as Northern analysis, quantitative RT PCR or RNase protection. Upon corroboration, the differentially expressed genes may be further characterized, and may be identified
25 as target and/or fingerprint genes, as discussed below.
30

Amplified sequences of differentially expressed genes obtained through, for example, differential display may be used to isolate full length clones of the corresponding gene. The full length coding portion of the gene may readily be isolated, without undue experimentation, by molecular biological techniques well known in the art. For example, the isolated differentially expressed amplified fragment may be labeled and used to screen a cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. As described above, the isolated, amplified gene fragments obtained through differential display have 5' terminal ends at some random point within the gene and usually have 3' terminal ends at a position corresponding to the 3' end of the transcribed portion of the gene. Once nucleotide sequence information from an amplified fragment is obtained, the remainder of the gene (*i.e.*, the 5' end of the gene, when utilizing differential display) may be obtained using, for example, RT-PCR.

In one embodiment of such a procedure for the identification and cloning of full length gene sequences, RNA may be isolated, following standard procedures, from an appropriate tissue or cellular source. A reverse transcription reaction may then be performed on the RNA using an oligonucleotide primer complimentary to the mRNA that corresponds to the amplified fragment, for the priming of first strand synthesis. Because the primer is anti-parallel to the mRNA, extension will proceed toward the 5' end of the mRNA. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Using the two primers, the 5' portion of the gene is amplified using PCR. Sequences obtained may then be isolated and recombined with previously isolated sequences to generate a full-length cDNA of the differentially expressed genes of the invention. For a review of cloning strategies and recombinant DNA techniques, see *e.g.*, Sambrook *et al.*, 1989, and Ausubel *et al.*, 1989.

Identification of Pathway Genes

Methods are described herein for the identification of pathway genes. "Pathway gene" refers to a gene whose gene product exhibits the ability to interact with gene products involved in pathogenic infections. A pathway gene may be differentially expressed and, therefore, may have the characteristics of a target and/or fingerprint gene.

Any method suitable for detecting protein-protein interactions may be employed for identifying pathway gene products by identifying interactions between gene products and gene products known to be involved in pathogenic infections. Such known gene products may be cellular or extracellular proteins. Those gene products which interact with such known gene products represent pathway gene products and the genes which encode them represent pathway genes.

Among the traditional methods which may be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Utilizing procedures such as these allows for the identification of pathway gene products. Once identified, a pathway gene product may be used, in conjunction with standard techniques, to identify its corresponding pathway gene. For example, at least a portion of the amino acid sequence of the pathway gene product may be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique (see, *e.g.*, Creighton, PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES, W. H. Freeman & Co., N.Y., pp.34-49, 1983). The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for pathway gene sequences. Screening may be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known. (see, *e.g.*, Ausubel, 1989, and Innis *et al.*, eds., PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, 1990, Academic Press, Inc., New York).

Methods may be employed which result in the simultaneous identification of pathway genes which encode the protein interacting with a protein involved in pathogenic infections. These methods include, for example, probing expression libraries with labeled protein known or suggested to be involved in such disorders, using this protein in a manner similar to the well known technique of antibody probing of λ gt11 libraries.

One method which detects protein interactions *in vivo*, the two-hybrid system, is described in detail for illustration only and not by way of limitation. One version of this system is been described in Chien *et al.*, 1991, *Proc. Natl. Acad. Sci. U.S.A.* 88, 9578-82, 1991, and is commercially available from Clontech (Palo Alto, Calif.). Briefly, utilizing such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to a known protein, in this case, a protein known to be involved in body weight disorders and or processes relevant to appetite and/or weight regulation, and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA which has been recombined into this plasmid as part of a cDNA library. The plasmids are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (*e.g.*, *lacZ*) whose regulatory region contains the transcription activator's binding sites. Either hybrid protein alone cannot activate transcription of the reporter gene: the DNA-binding domain hybrid cannot because it does not provide activation function and the activation domain hybrid cannot because it cannot localize to the activator's binding sites. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

The two-hybrid system or related methodology may be used to screen activation domain libraries for proteins that interact with a known "bait" gene product. By way of example, and not by way of limitation, gene products known to be involved in body weight disorders and/or appetite or body weight regulation may be used as the

bait gene products. These include but are not limited to the intracellular domain of receptors for such hormones as neuropeptide Y, galanin, interostatin, insulin, and CCK. Total genomic or cDNA sequences are fused to the DNA encoding an activation domain. This library and a plasmid encoding a hybrid of the bait gene product fused to the DNA-binding domain are cotransformed into a yeast reporter strain, and the resulting transformants are screened for those that express the reporter gene. For example, and not by way of limitation, the bait gene can be cloned into a vector such that it is translationally fused to the DNA encoding the DNA-binding domain of the GAL4 protein. These colonies are purified and the library plasmids responsible for reporter gene expression are isolated. DNA sequencing is then used to identify the proteins encoded by the library plasmids.

A cDNA library of the cell line from which proteins that interact with bait gene product are to be detected can be made using methods routinely practiced in the art. According to the particular system described herein, for example, the cDNA fragments can be inserted into a vector such that they are translationally fused to the activation domain of GAL4. This library can be co-transformed along with the bait gene-GAL4 fusion plasmid into a yeast strain which contains a lacZ gene driven by a promoter which contains GAL4 activation sequence. A cDNA encoded protein, fused to GAL4 activation domain, that interacts with bait gene product will reconstitute an active GAL4 protein and thereby drive expression of the lacZ gene. Colonies which express lacZ can be detected by their blue color in the presence of X-gal. The cDNA can then be purified from these strains, and used to produce and isolate the bait gene-interacting protein using techniques routinely practiced in the art. Once a pathway gene has been identified and isolated, it may be further characterized, as described below..

Characterization of Differentially Expressed and Pathway Genes

Differentially expressed and pathway genes, such as those identified via the methods discussed above, as well as genes identified by alternative means, may be further

characterized by utilizing, for example, methods such as those discussed herein. Such genes will be referred to herein as "identified genes." Analyses such as those described herein, yield information regarding the biological function of the identified genes. An assessment of the biological function of the differentially expressed genes, in addition, will allow for their designation as target and/or fingerprint genes.

Specifically, any of the differentially expressed genes whose further characterization indicates that a modulation of the gene's expression or a modulation of the gene product's activity may ameliorate any of the body weight disorders of interest will be designated "target genes," as defined above. Such target genes and target gene products, along with those discussed below, will constitute the focus of the compound discovery strategies discussed below. Further, such target genes, target gene products and/or modulating compounds can be used as part of the treatment methods described below.

Any of the differentially expressed genes whose further characterization indicates that such modulations may not positively affect body weight disorders of interest, but whose expression pattern contributes to a gene expression "fingerprint" pattern correlative of, for example, a malignant state will be designated a "fingerprint gene." It should be noted that each of the target genes may also function as fingerprint genes, as well as may all or a portion of the pathway genes.

Pathway genes may also be characterized according to techniques such as those described herein. Those pathway genes which yield information indicating that they are differentially expressed and that modulation of the gene's expression or a modulation of the gene product's activity may ameliorate any of the disorders of interest will be also be designated "target genes." Such target genes and target gene products, along with those discussed above, will constitute the focus of the compound discovery strategies discussed below and can be used as part of treatment methods.

Characterization of one or more of the pathway genes may reveal a lack of differential expression, but evidence that modulation of the gene's activity or expression may, nonetheless, ameliorate body weight disorder symptoms. In such cases, these genes and gene products would also be considered a focus of the compound discovery strategies. In instances wherein a pathway gene's characterization indicates that modulation of gene expression or gene product activity may not positively affect body weight disorders of interest, but whose expression is differentially expressed and contributes to a gene expression fingerprint pattern correlative of, for example, a pathogenic infection, such pathway genes may additionally be designated as fingerprint genes.

A variety of techniques can be utilized to further characterize the identified genes. First, the nucleotide sequence of the identified genes, which may be obtained by utilizing standard techniques well known to those of skill in the art, may, for example, be used to reveal homologies to one or more known sequence motifs which may yield information regarding the biological function of the identified gene product.

Second, an analysis of the tissue and/or cell type distribution of the mRNA produced by the identified genes may be conducted, utilizing standard techniques well known to those of skill in the art. Such techniques may include, for example, Northern, RNase protection and RT-PCR analyses. Such analyses provide information as to, for example, whether the identified genes are expressed in tissues or cell types expected to contribute to the disorders of interest. Such analyses may also provide quantitative information regarding steady state mRNA regulation, yielding data concerning which of the identified genes exhibits a high level of regulation in, preferably, tissues which may be expected to contribute to the disorders of interest. Additionally, standard *in situ* hybridization techniques may be utilized to provide information regarding which cells within a given tissue express the identified gene. Such an analysis may provide information regarding the biological function of an identified gene relative to a given

disorder in instances wherein only a subset of the cells within the tissue is thought to be relevant to the body weight disorder.

Third, the sequences of the identified genes may be used, utilizing standard techniques, to place the genes onto genetic maps, *e.g.*, mouse (Copeland and Jenkins, *Trends in Genetics* 7, 113-18, 1991) and human genetic maps (Cohen *et al.*, *Nature* 366, 698-701, 1993). Such mapping information may yield information regarding the genes' importance to human disease by, for example, identifying genes which map within genetic regions to which known genetic disorders map.

Fourth, the biological function of the identified genes may be more directly assessed by utilizing relevant *in vivo* and *in vitro* systems. *In vivo* systems may include, but are not limited to, animal systems which naturally exhibit body weight disorder-like symptoms, or ones which have been engineered to exhibit such symptoms. Further, such systems may include systems for the further characterization of body weight disorders, and/or appetite or body weight regulation, and may include, but are not limited to, naturally occurring and transgenic animal systems. *In vitro* systems may include, but are not limited to, cell-based systems comprising cell types known or suspected of contributing to the body weight disorder of interest. Such cells may be wild type cells, or may be non-wild type cells containing modifications known to, or suspected of, contributing to the body weight disorder of interest.

In further characterizing the biological function of the identified genes, the expression of these genes may be modulated within the *in vivo* and/or *in vitro* systems, *i.e.*, either overexpressed or underexpressed in, for example, transgenic animals and/or cell lines, and its subsequent effect on the system then assayed. Alternatively, the activity of the product of the identified gene may be modulated by either increasing or decreasing the level of activity in the *in vivo* and/or *in vitro* system of interest, and its subsequent effect then assayed.

The information obtained through such characterizations may suggest relevant methods for the treatment of disorders involving the gene of interest. Further, relevant methods for the treatment of such disorders involving the gene of interest may be suggested by information obtained from such characterizations. For example, treatment may include a modulation of gene expression and/or gene product activity. Characterization procedures such as those described herein may indicate where such modulation should involve an increase or a decrease in the expression or activity of the gene or gene product of interest.

10 Screening Methods

The invention provides methods for identifying modulators, *i.e.*, candidate or test compounds which bind to MASP-like serine protease polypeptides or polynucleotides and/or have a stimulatory or inhibitory effect on, for example, expression or activity of the MASP-like serine protease polypeptide or polynucleotide, so as to regulate degradation of the extracellular matrix. Decreased extracellular matrix degradation is useful for preventing or suppressing malignant cells from metastasizing. Increased extracellular matrix degradation may be desired, for example, in developmental disorders characterized by inappropriately low levels of extracellular matrix degradation or in regeneration.

The invention provides assays for screening test compounds which bind to or modulate the activity of a MASP-like serine protease polypeptide or a MASP-like serine protease polynucleotide. A test compound preferably binds to a MASP-like serine protease polypeptide or polynucleotide. More preferably, a test compound decreases a MASP-like serine protease activity of a MASP-like serine protease polypeptide or expression of a MASP-like serine protease polynucleotide by at least about 10, preferably about 50, more preferably about 75, 90, or 100% relative to the absence of the test compound.

30

Test Compounds

Test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity. Such compounds also may include, but are not limited to, other cellular proteins, peptides such as, for example, soluble peptides, including but not limited to, Ig-tailed fusion peptides, comprising extracellular portions of target gene product transmembrane receptors, and members of random peptide libraries (Lam, *et al.*, *Nature* 354, 82-84, 1991; Houghten *et al.*, *Nature* 354, 84-86, 1991), made of D- and/or L-configuration amino acids, phosphopeptides (including, but not limited to members of random or partially degenerate, directed phosphopeptide libraries (Songyang *et al.*, *Cell* 72, 767-78, 1993), antibodies (including, but not limited to, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, F(ab')₂ and FAb expression library fragments, and epitope-binding fragments thereof), and small organic or inorganic molecules.

The compounds can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, and can be produced recombinantly, or synthesized by chemical methods known in the art. If desired, test compounds can be obtained using any of the numerous combinatorial library methods known in the art, including but not limited to, biological libraries, spatially addressable parallel solid phase or solution phase libraries, synthetic library methods requiring deconvolution, the "one-bead one-compound" library method, and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer, or small molecule libraries of compounds. See Lam, *Anticancer Drug Des.* 12, 145, 1997.

Methods for the synthesis of molecular libraries are well known in the art (see, for example, DeWitt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 90, 6909, 1993; Erb *et al.* *Proc. Natl. Acad. Sci. U.S.A.* 91, 11422, 1994; Zuckermann *et al.*, *J. Med. Chem.* 37, 2678,

1994; Cho *et al.*, *Science* 261, 1303, 1993; Carell *et al.*, *Angew. Chem. Int. Ed. Engl.* 33, 2059, 1994; Carell *et al.*, *Angew. Chem. Int. Ed. Engl.* 33, 2061; Gallop *et al.*, *J. Med. Chem.* 37, 1233, 1994). Libraries of compounds can be presented in solution (see, e.g., Houghten, *Biotechniques* 13, 412-421, 1992), or on beads (Lam, *Nature* 354, 82-84, 1991), chips (Fodor, *Nature* 364, 555-556, 1993), bacteria or spores (Ladner, U.S. Patent 5,223,409), plasmids (Cull *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89, 1865-1869, 1992), or phage (Scott & Smith, *Science* 249, 386-390, 1990; Devlin, *Science* 249, 404-406, 1990); Cwirla *et al.*, *Proc. Natl. Acad. Sci.* 97, 6378-6382, 1990; Felici, *J. Mol. Biol.* 222, 301-310, 1991; and Ladner, U.S. Patent 5,223,409).

High Throughput Screening

Test compounds can be screened for the ability to bind to MASP-like serine protease polypeptides or polynucleotides or to affect MASP-like serine protease activity or MASP-like serine protease gene expression using high throughput screening. Using high throughput screening, many discrete compounds can be tested in parallel so that large numbers of test compounds can be quickly screened. The most widely established techniques utilize 96-well microtiter plates. The wells of the microtiter plates typically require assay volumes that range from 50 to 500 μ l. In addition to the plates, many instruments, materials, pipettors, robotics, plate washers, and plate readers are commercially available to fit the 96-well format.

Alternatively, "free format assays," or assays that have no physical barrier between samples, can be used. For example, an assay using pigment cells (melanocytes) in a simple homogeneous assay for combinatorial peptide libraries is described by Jayawickreme *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 19, 1614-18 (1994). The cells are placed under agarose in petri dishes, then beads that carry combinatorial compounds are placed on the surface of the agarose. The combinatorial compounds are partially released the compounds from the beads. Active compounds can be visualized as dark pigment areas because, as the compounds diffuse locally into the gel matrix, the active compounds cause the cells to change colors.

Another example of a free format assay is described by Chelsky, "Strategies for Screening Combinatorial Libraries: Novel and Traditional Approaches," reported at the First Annual Conference of The Society for Biomolecular Screening in Philadelphia, Pa. (Nov. 7-10, 1995). Chelsky placed a simple homogenous enzyme assay for carbonic anhydrase inside an agarose gel such that the enzyme in the gel would cause a color change throughout the gel. Thereafter, beads carrying combinatorial compounds via a photolinker were placed inside the gel and the compounds were partially released by UV-light. Compounds that inhibited the enzyme were observed as local zones of inhibition having less color change.

Yet another example is described by Salmon *et al.*, *Molecular Diversity* 2, 57-63 (1996). In this example, combinatorial libraries were screened for compounds that had cytotoxic effects on cancer cells growing in agar.

Another high throughput screening method is described in Beutel *et al.*, U.S. Patent 5,976,813. In this method, test samples are placed in a porous matrix. One or more assay components are then placed within, on top of, or at the bottom of a matrix such as a gel, a plastic sheet, a filter, or other form of easily manipulated solid support. When samples are introduced to the porous matrix they diffuse sufficiently slowly, such that the assays can be performed without the test samples running together.

Binding Assays

For binding assays, the test compound is preferably a small molecule which binds to and occupies the active site or a fibronectin domain of the MASP-like serine protease polypeptide, thereby making the active site or fibronectin domain inaccessible to substrate such that normal biological activity is prevented. Examples of such small molecules include, but are not limited to, small peptides or peptide-like molecules. In binding assays, either the test compound or the MASP-like serine protease polypeptide can comprise a detectable label, such as a fluorescent, radioisotopic,

chemiluminescent, or enzymatic label, such as horseradish peroxidase, alkaline phosphatase, or luciferase. Detection of a test compound which is bound to the MASP-like serine protease polypeptide can then be accomplished, for example, by direct counting of radioemmission, by scintillation counting, or by determining conversion of an appropriate substrate to a detectable product.

Alternatively, binding of a test compound to a MASP-like serine protease polypeptide can be determined without labeling either of the interactants. For example, a microphysiometer can be used to detect binding of a test compound with a target polypeptide. A microphysiometer (*e.g.*, Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a test compound and a MASP-like-serine protease polypeptide. (McConnell *et al.*, *Science* 257, 1906-1912, 1992).

Determining the ability of a test compound to bind to a MASP-like serine protease polypeptide also can be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander & Urbaniczky, *Anal. Chem.* 63, 2338-2345, 1991, and Szabo *et al.*, *Curr. Opin. Struct. Biol.* 5, 699-705, 1995. BIA is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In yet another aspect of the invention, a MASP-like serine protease polypeptide can be used as a "bait protein" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent 5,283,317; Zervos *et al.*, *Cell* 72, 223-232, 1993; Madura *et al.*, *J. Biol. Chem.* 268, 12046-12054, 1993; Bartel *et al.*, *Biotechniques* 14, 920-924, 1993; Iwabuchi *et al.*, *Oncogene* 8, 1693-1696, 1993; and Brent W094/10300), to identify other proteins which bind to or interact with the MASP-like serine protease polypeptide and modulate its activity.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. For example, in one construct a polynucleotide encoding a MASP-like serine protease polypeptide is fused to a polynucleotide encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence that encodes an unidentified protein ("prey" or "sample") is fused to a polynucleotide that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact *in vivo* to form an protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ), which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the DNA sequence encoding the protein which interacts with the MASP-like serine protease polypeptide.

It may be desirable to immobilize either the MASP-like serine protease polypeptide (or polynucleotide) or the test compound to facilitate separation of bound from unbound forms of one or both of the interactants, as well as to accommodate automation of the assay. Thus, either the MASP-like serine protease polypeptide (or polynucleotide) or the test compound can be bound to a solid support. Suitable solid supports include, but are not limited to, glass or plastic slides, tissue culture plates, microtiter wells, tubes, silicon chips, or particles such as beads (including, but not limited to, latex, polystyrene, or glass beads). Any method known in the art can be used to attach the MASP-like serine protease polypeptide (or polynucleotide) or test compound to a solid support, including use of covalent and non-covalent linkages, passive absorption, or pairs of binding moieties attached respectively to the polypeptide or test compound and the solid support. Test compounds are preferably bound to the solid support in an array, so that the location of individual test compounds can be tracked. Binding of a test compound to a MASP-like serine

protease polypeptide (or polynucleotide) can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and microcentrifuge tubes.

5 In one embodiment, a MASP-like serine protease polypeptide is a fusion protein comprising a domain that allows the MASP-like serine protease polypeptide to be bound to a solid support. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or
10 glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and the non-adsorbed MASP-like serine protease polypeptide; the mixture is then incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components. Binding of the interactants can be determined either directly or indirectly, as
15 described above. Alternatively, the complexes can be dissociated from the solid support before binding is determined.

Other techniques for immobilizing polypeptides or polynucleotides on a solid support also can be used in the screening assays of the invention. For example, either
20 a MASP-like serine protease polypeptide (or polynucleotide) or a test compound can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated MASP-like serine protease polypeptides or test compounds can be prepared from biotin-NHS(N-hydroxysuccinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.) and immobilized in the wells of
25 streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies which specifically bind to a MASP-like serine protease polypeptide polynucleotides, or a test compound, but which do not interfere with a desired binding site, such as the active site or a fibronectin domain of the MASP-like serine protease polypeptide, can be derivatized to the wells of the plate. Unbound target or protein can be trapped
30 in the wells by antibody conjugation.

- 53 -

Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies which specifically bind to the MASP-like serine protease polypeptide (or polynucleotides) or test compound, enzyme-linked assays which rely on detecting a MASP-like serine protease activity of the MASP-like serine protease polypeptide, and SDS gel electrophoresis under non-reducing conditions.

Screening for test compounds which bind to a MASP-like serine protease polypeptide or polynucleotide also can be carried out in an intact cell. Any cell which comprises a MASP-like serine protease polynucleotide or polypeptide can be used in a cell-based assay system. A MASP-like serine protease polynucleotide can be naturally occurring in the cell or can be introduced using techniques such as those described above. Either a primary culture or an established cell line, including neoplastic cell lines such as the colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, and the H392 glioblastoma cell line, can be used. An intact cell is contacted with a test compound. Binding of the test compound to a MASP-like serine protease polypeptide or polynucleotide is determined as described above, after lysing the cell to release the MASP-like serine protease polypeptide-test compound complex.

Enzyme Assays

Test compounds can be tested for the ability to increase or decrease a MASP-like serine protease activity of a MASP-like serine protease polypeptide. MASP-like serine protease activity can be measured, for example, as described in Matsushita *et al.*, 2000. MASP-like serine protease activity can be measured after contacting either a purified MASP-like serine protease polypeptide, a cell extract, or an intact cell with a test compound. A test compound which decreases MASP-like serine protease activity by at least about 10, preferably about 50, more preferably about 75, 90, or 100% is identified as a potential therapeutic agent for decreasing extracellular matrix

degradation. A test compound which increases MASP-like serine protease activity by at least about 10, preferably about 50, more preferably about 75, 90, or 100% is identified as a potential therapeutic agent for increasing extracellular matrix degradation.

5

Gene Expression

In another embodiment, test compounds which increase or decrease MASP-like serine protease gene expression are identified. A MASP-like serine protease polynucleotide is contacted with a test compound, and the expression of an RNA or polypeptide product of the MASP-like serine protease polynucleotide is determined. The level of expression of MASP-like serine protease mRNA or polypeptide in the presence of the test compound is compared to the level of expression of MASP-like serine protease mRNA or polypeptide in the absence of the test compound. The test compound can then be identified as a modulator of expression based on this comparison. For example, when expression of MASP-like serine protease mRNA or polypeptide is greater in the presence of the test compound than in its absence, the test compound is identified as a stimulator or enhancer of MASP-like serine protease mRNA or polypeptide expression. Alternatively, when expression of the mRNA or protein is less in the presence of the test compound than in its absence, the test compound is identified as an inhibitor of MASP-like serine protease mRNA or polypeptide expression.

The level of MASP-like serine protease mRNA or polypeptide expression in the cells can be determined by methods well known in the art for detecting mRNA or protein. Either qualitative or quantitative methods can be used. The presence of polypeptide products of a MASP-like serine protease polynucleotide can be determined, for example, using a variety of techniques known in the art, including immunochemical methods such as radioimmunoassay, Western blotting, and immunohistochemistry. Alternatively, polypeptide synthesis can be determined in

vivo, in a cell culture, or in an *in vitro* translation system by detecting incorporation of labeled amino acids into a MASP-like serine protease polypeptide.

Such screening can be carried out either in a cell-free assay system or in an intact
5 cell. Any cell which expresses a MASP-like serine protease polynucleotide can be
used in a cell-based assay system. The MASP-like serine protease polynucleotide
can be naturally occurring in the cell or can be introduced using techniques such as
those described above. Either a primary culture or an established cell line, including
neoplastic cell lines such as the colon cancer cell lines HCT116, DLD1, HT29,
10 Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-
468, SK-BR3, and BT-474, the A549 lung cancer cell line, and the H392
glioblastoma cell line, can be used.

Pharmaceutical Compositions

15 The invention also provides pharmaceutical compositions which can be administered
to a patient to achieve a therapeutic effect. Pharmaceutical compositions of the
invention can comprise a MASP-like serine protease polypeptide, MASP-like serine
protease polynucleotide, antibodies which specifically bind to a MASP-like serine
20 protease polypeptide, or mimetics, agonists, antagonists, or inhibitors of a MASP-
like serine protease polypeptide. The compositions can be administered alone or in
combination with at least one other agent, such as stabilizing compound, which can
be administered in any sterile, biocompatible pharmaceutical carrier, including, but
not limited to, saline, buffered saline, dextrose, and water. The compositions can be
25 administered to a patient alone, or in combination with other agents, drugs or
hormones.

In addition to the active ingredients, these pharmaceutical compositions can contain
suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries
30 which facilitate processing of the active compounds into preparations which can be
used pharmaceutically. Pharmaceutical compositions of the invention can be ad-

ministered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, parenteral, topical, sublingual, or rectal means. Pharmaceutical compositions for oral administration can be formulated using
5 pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

10 Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn,
15 wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

20 Dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which also can contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments can be added to
25 the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as
30 glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium

- 57 -

stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

5 Pharmaceutical formulations suitable for parenteral administration can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions can contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally,
10 suspensions of the active compounds can be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers also can be used for delivery. Optionally, the suspension also can contain suitable stabilizers or agents which increase the
15 solubility of the compounds to allow for the preparation of highly concentrated solutions. For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

20 The pharmaceutical compositions of the present invention can be manufactured in a manner that is known in the art, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. The pharmaceutical composition can be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric,
25 sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation can be a lyophilized powder which can contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

30

Further details on techniques for formulation and administration can be found in the latest edition of REMINGTON'S PHARMACEUTICAL SCIENCES (Maack Publishing Co., Easton, Pa.). After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated
5 condition. Such labeling would include amount, frequency, and method of administration.

Therapeutic Indications and Methods

10 In humans, binding of MASP to carbohydrates initiates the lectin pathway, which is involved in host defense against pathogens. Takahashi *et al.*, 1999; Endo *et al.*, 1998. Thus, MASP-like serine protease or reagents which increase its activity can be administered to patients to provide a defense against pathogens, such as viruses, bacteria, mycoplasma, fungi, protozoa, helminths, rickettsia, chloamydiae, parasites,
15 prions, and the like. Alternatively, in diseases in which MASP-like serine protease is overexpressed, it may be useful to decrease expression levels using reagents which specifically bind to MASP-like serine protease or its gene, including antibodies, ribozymes, antisense oligonucleotides, or other molecules.

20 The invention further pertains to the use of novel agents identified by the screening assays described above. Accordingly, it is within the scope of this invention to use a test compound identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a modulating agent, an antisense nucleic acid molecule, a specific antibody, ribozyme, or a polypeptide-binding
25 partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

A reagent which affects MASP-like serine protease activity can be administered to a human cell, either *in vitro* or *in vivo*, to reduce MASP-like serine protease activity. The reagent preferably binds to an expression product of a human MASP-like serine protease gene. If the expression product is a polypeptide, the reagent is preferably an antibody. For treatment of human cells *ex vivo*, an antibody can be added to a preparation of stem cells which have been removed from the body. The cells can then be replaced in the same or another human body, with or without clonal propagation, as is known in the art.

10 In one embodiment, the reagent is delivered using a liposome. Preferably, the liposome is stable in the animal into which it has been administered for at least about 30 minutes, more preferably for at least about 1 hour, and even more preferably for at least about 24 hours. A liposome comprises a lipid composition that is capable of targeting a reagent, particularly a polynucleotide, to a particular site in an animal, such as a human. Preferably, the lipid composition of the liposome is capable of targeting to a specific organ of an animal, such as the lung or liver.

A liposome useful in the present invention comprises a lipid composition that is capable of fusing with the plasma membrane of the targeted cell to deliver its contents to the cell. Preferably, the transfection efficiency of a liposome is about 0.5 μg of DNA per 16 nmole of liposome delivered to about 10^6 cells, more preferably about 1.0 μg of DNA per 16 nmol of liposome delivered to about 10^6 cells, and even more preferably about 2.0 μg of DNA per 16 nmol of liposome delivered to about 10^6 cells. Preferably, a liposome is between about 100 and 500 nm, more preferably between about 150 and 450 nm, and even more preferably between about 200 and 400 nm in diameter.

Suitable liposomes for use in the present invention include those liposomes standardly used in, for example, gene delivery methods known to those of skill in the art. More preferred liposomes include liposomes having a polycationic lipid composition and/or liposomes having a cholesterol backbone conjugated to

- 60 -

polyethylene glycol. Optionally, a liposome comprises a compound capable of targeting the liposome to a tumor cell, such as a tumor cell ligand exposed on the outer surface of the liposome.

5 Complexing a liposome with a reagent such as an antisense oligonucleotide or ribozyme can be achieved using methods which are standard in the art (see, for example, U.S. Patent 5,705,151). Preferably, from about 0.1 μ g to about 10 μ g of polynucleotide is combined with about 8 nmol of liposomes, more preferably from about 0.5 μ g to about 5 μ g of polynucleotides are combined with about 8 nmol liposomes, and even more preferably about 1.0 μ g of polynucleotides is combined with
10 about 8 nmol liposomes.

In another embodiment, antibodies can be delivered to specific tissues *in vivo* using receptor-mediated targeted delivery. Receptor-mediated DNA delivery techniques
15 are taught in, for example, Findeis *et al.* *Trends in Biotechnol.* 11, 202-05 (1993); Chiou *et al.*, GENE THERAPEUTICS: METHODS AND APPLICATIONS OF DIRECT GENE TRANSFER (J.A. Wolff, ed.) (1994); Wu & Wu, *J. Biol. Chem.* 263, 621-24 (1988); Wu *et al.*, *J. Biol. Chem.* 269, 542-46 (1994); Zenke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 87, 3655-59 (1990); Wu *et al.*, *J. Biol. Chem.* 266, 338-42 (1991).
20

If the reagent is a single-chain antibody, polynucleotides encoding the antibody can be constructed and introduced into a cell either *ex vivo* or *in vivo* using well-established techniques including, but not limited to, transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads,
25 protoplast fusion, viral infection, electroporation, "gene gun," and DEAE- or calcium phosphate-mediated transfection.

Determination of a Therapeutically Effective Dose

The determination of a therapeutically effective dose is well within the capability of those skilled in the art. A therapeutically effective dose refers to that amount of active ingredient which increases or decreases extracellular matrix degradation relative to that which occurs in the absence of the therapeutically effective dose.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in animal models, usually mice, rabbits, dogs, or pigs. The animal model also can be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

Therapeutic efficacy and toxicity, e.g., ED_{50} (the dose therapeutically effective in 50% of the population) and LD_{50} (the dose lethal to 50% of the population), can be determined by standard pharmaceutical procedures in cell cultures or experimental animals. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the ratio, LD_{50}/ED_{50} .

Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active ingredient or to maintain the desired effect. Factors which can be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and

frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on the half-life and clearance rate of the particular formulation.

5

Normal dosage amounts can vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

10

Effective *in vivo* dosages of an antibody are in the range of about 5 μ g to about 50 μ g/kg, about 50 μ g to about 5 mg/kg, about 100 μ g to about 500 μ g/kg of patient body weight, and about 200 to about 250 μ g/kg of patient body weight. For administration of polynucleotides encoding single-chain antibodies, effective *in vivo* dosages are in the range of about 100 ng to about 200 ng, 500 ng to about 50 mg, about 1 μ g to about 2 mg, about 5 μ g to about 500 μ g, and about 20 μ g to about 100 μ g of DNA.

15

20

If the expression product is mRNA, the reagent is preferably an antisense oligonucleotide or a ribozyme. Polynucleotides which express antisense oligonucleotides or ribozymes can be introduced into cells by a variety of methods, as described above.

25

Preferably, a reagent reduces expression of a MASP-like serine protease polynucleotide or activity of a MASP-like serine protease polypeptide by at least about 10, preferably about 50, more preferably about 75, 90, or 100% relative to the absence of the reagent. The effectiveness of the mechanism chosen to decrease the level of expression of a MASP-like serine protease polynucleotide or the activity of a

30

MASP-like serine protease polypeptide can be assessed using methods well known in the art, such as hybridization of nucleotide probes to MASP-like serine protease-specific mRNA, quantitative RT-PCR, immunologic detection of a MASP-like serine protease polypeptide, or measurement of MASP-like serine protease activity.

5

In any of the embodiments described above, any of the pharmaceutical compositions of the invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy can be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents can act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

10

Any of the therapeutic methods described above can be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

15

The above disclosure generally describes the present invention, and all patents and patent applications cited in this disclosure are expressly incorporated herein. A more complete understanding can be obtained by reference to the following specific examples which are provided for purposes of illustration only and are not intended to limit the scope of the invention.

20

EXAMPLE 1

Detection of MASP-like serine protease activity

5 The polynucleotide of SEQ ID NO: 1 is inserted into the expression vector pCEV4 and the expression vector pCEV4-MASP-like serine protease polypeptide obtained is transfected into human embryonic kidney 293 cells.

10 Protease activity of cellular extracts from the transfected cells are measured using thiobenzylester substrates, as described in U.S. Patent 5,500,344. For monitoring enzyme activities from granules and column fractions, assays are performed at room temperature using 0.5 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) (Sigma) to detect the HSBzl leaving group ($\epsilon_{410} = 13600 \text{ M}^{-1} \text{ cm}^{-1}$).

15 Furthermore, BLT-esterase activity is estimated using a microtiter assay (Green and Shaw, *Anal. Biochem.* 93, 223-226, 1979). Briefly, 50 μl of sample is added to 100 μl of 1 mM DTNB, made up in 10 mM HEPES, 1 mM CaCl_2 , 1 mM MgCl_2 , pH 7.2. The reaction is initiated by the addition of 50 μl of BLT (Sigma) to give a final concentration of 500 μM . For Metase determinations, 50 μl of dilutions of the sample in 0.1 M HEPES, 0.05 M CaCl_2 , pH 7.5, are added to 100 μl of 1 mM DTNB, and the reaction is initiated by the addition of 50 μl of Boc-Ala-Ala-Met-S Benzyl (Bzl) to give a final concentration of 150 μM . The duration of the assay depends on color development, the rate of which is measured (O.D._{410}) on a Dynatech MR 5000 microplate reader. Controls of sample and DTNB alone or DTNB and substrate alone are run.

25 Additionally, peptide thiobenzyl ester substrates are used to measure protease activities. The chymase substrate Suc-Phe-Leu-Phe-SBzl is purchased from BACHEM Bioscience Inc., Philadelphia, Pa. Z-Arg-SBzl (the tryptase substrate, Kam *et al.*, *J. Biol. Chem.* 262, 3444-3451, 1987); Boc-Ala-Ala-AA-SBzl (AA=Asp, 30 Met, Leu, Nle, or Ser), and Suc-Ala-Ala-Met-SBzl (Otake *et al.*, *Biochemistry* 30, 2217-2227, 1991); Harper *et al.*, *Biochemistry* 23, 2995-3002, 1984) are synthesized

previously. Boc-Ala-Ala-Asp-SBzl is the substrate for Asp-ase and peptide thiobenzyl esters containing Met, Leu or Nle are substrates for Met-ase SP. Assays are performed at room temperature in 0.1 M, HEPES buffer, pH 7.5, containing 0.01 M CaCl_2 and 8% Me_2O using 0.34 mM 4,4'-dithiodipyridine (Aldrithiol-4, Aldrich Chemical Co., Milwaukee, Wis.) to detect HSBzl leaving group that reacts with 4,4'-dithiodipyridine to release thiopyridone ($\epsilon_{324}=19800 \text{ M}^{-1} \text{ cm}^{-1}$, Grasetti and Murray, *Arch. Biochem. Biophys.* 119, 41-49, 1967). The initial rates are measured at 324 nm using a Beckman 35 spectrophotometer when 10-25 μl of an enzyme stock solution is added to a cuvette containing 2.0 ml of buffer, 150 μl of 4,4'-dithiodipyridine, and 25 μl of substrate. The same volume of substrate and 4,4'-dithiodipyridine are added to the reference cell in order to compensate for the background hydrolysis rate of the substrates. Initial rates are measured in duplicate for each substrate concentration and are averaged in each case. Substrate concentrations are 100-133 μM . The MASP-like serine protease activity of the polypeptide of SEQ ID NO: 2 is shown.

EXAMPLE 2

Identification of a test compound which binds to a MASP-like serine protease polypeptide

Purified MASP-like serine protease polypeptides comprising a glutathione-S-transferase protein and absorbed onto glutathione-derivatized wells of 96-well microtiter plates are contacted with test compounds from a small molecule library at pH 7.0 in a physiological buffer solution. MASP-like serine protease polypeptides comprise the amino acid sequence shown in SEQ ID NO:2. The test compounds comprise a fluorescent tag. The samples are incubated for 5 minutes to one hour. Control samples are incubated in the absence of a test compound.

The buffer solution containing the test compounds is washed from the wells. Binding of a test compound to a MASP-like serine protease polypeptide is detected by fluorescence measurements of the contents of the wells. A test compound which increases the fluorescence in a well by at least 15% relative to fluorescence of a well

in which a test compound was not incubated is identified as a compound which binds to a MASP-like serine protease polypeptide.

EXAMPLE 3

5 Identification of a test compound which decreases MASP-like serine protease activity

Cellular extracts from the human colon cancer cell line HCT116 are contacted with test compounds from a small molecule library and assayed for MASP-like serine protease activity. Control extracts, in the absence of a test compound, also are
10 assayed. Matriptase activity is measured as described in Matsushita *et al.*, 2000. A test compound which decreases MASP-like serine protease activity of the extract relative to the control extract by at least 20% is identified as a MASP-like serine protease inhibitor.

EXAMPLE 4

15 Identification of a test compound which decreases MASP-like serine protease gene expression

A test compound is administered to a culture of the breast tumor cell line MDA-468
20 and incubated at 37 °C for 10 to 45 minutes. A culture of the same type of cells incubated for the same time without the test compound provides a negative control.

RNA is isolated from the two cultures as described in Chirgwin *et al.*, *Biochem. 18*, 5294-99, 1979). Northern blots are prepared using 20 to 30 µg total RNA and
25 hybridized with a ³²P-labeled MASP-like serine protease-specific probe at 65 °C in Express-hyb (CLONTECH). The probe comprises at least 11 contiguous nucleotides selected from the complement of SEQ ID NO:1. A test compound which decreases the MASP-like serine protease -specific signal relative to the signal obtained in the absence of the test compound is identified as an inhibitor of MASP-like serine
30 protease gene expression.

EXAMPLE 5*Treatment of a bacterial infection with MASP-like serine protease*

- 5 An expression construct which expresses MASP-like serine protease is administered to a patient with a bacterial infection. The patient is now capable of producing higher levels of MASP-like serine protease. The patient's bacterial infection is monitored over a period of days or weeks. Additional injections of the expression construct can be given during that time. The severity of the patient's bacterial infection is decreased due to increased MASP-like serine protease activity.

CLAIMS

1. An isolated polynucleotide encoding a MASP-like serine protease polypeptide and being selected from the group consisting of:
 - 5 a) a polynucleotide encoding a MASP-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to
the amino acid sequence shown in SEQ ID NO: 2; and
10 the amino acid sequence shown in SEQ ID NO: 2.
 - b) a polynucleotide comprising the sequence of SEQ ID NO: 1;
 - c) a polynucleotide which hybridizes under stringent conditions to a polynucleotide specified in (a) and (b);
 - d) a polynucleotide the sequence of which deviates from the polynucleotide sequences specified in (a) to (c) due to the degeneration of
15 the genetic code; and
 - e) a polynucleotide which represents a fragment, derivative or allelic variation of a polynucleotide sequence specified in (a) to (d).
- 20 2. An expression vector containing any polynucleotide of claim 1.
3. A host cell containing the expression vector of claim 2.
4. A substantially purified MASP-like serine protease polypeptide encoded by a
25 polynucleotide of claim 1.
5. A method for producing a MASP-like serine protease polypeptide, wherein the method comprises the following steps:
 - a) culturing the host cell of claim 3 under conditions suitable for the
30 expression of the MASP-like serine protease polypeptide; and

- b) recovering the MASP-like serine protease polypeptide from the host cell culture.
- 5 6. A method for detection of a polynucleotide encoding a MASP-like serine protease polypeptide in a biological sample comprising the following steps:
- a) hybridizing any polynucleotide of claim 1 to a nucleic acid material of a biological sample, thereby forming a hybridization complex; and
- b) detecting said hybridization complex.
- 10 7. The method of claim 6, wherein before hybridization, the nucleic acid material of the biological sample is amplified.
8. A method for the detection of a polynucleotide of claim 1 or a MASP-like serine protease polypeptide of claim 4 comprising the steps of:
- 15 contacting a biological sample with a reagent which specifically interacts with the polynucleotide or the MASP-like serine protease polypeptide.
9. A diagnostic kit for conducting the method of any one of claims 6 to 8.
- 20 10. A method of screening for agents which decrease the activity of a MASP-like serine protease, comprising the steps of:
- contacting a test compound with any MASP-like serine protease polypeptide encoded by any polynucleotide of claim 1;
- 25 detecting binding of the test compound to the MASP-like serine protease polypeptide, wherein a test compound which binds to the polypeptide is identified as a potential therapeutic agent for decreasing the activity of a MASP-like serine protease.
- 30 11. A method of screening for agents which regulate the activity of a MASP-like serine protease, comprising the steps of:

contacting a test compound with a MASP-like serine protease polypeptide encoded by any polynucleotide of claim 1; and

detecting a MASP-like serine protease activity of the polypeptide, wherein a test compound which increases the MASP-like serine protease activity is identified as a potential therapeutic agent for increasing the activity of the MASP-like serine protease, and wherein a test compound which decreases the MASP-like serine protease activity of the polypeptide is identified as a potential therapeutic agent for decreasing the activity of the MASP-like serine protease.

12. A method of screening for agents which decrease the activity of a MASP-like serine protease, comprising the steps of:

contacting a test compound with any polynucleotide of claim 1 and detecting binding of the test compound to the polynucleotide, wherein a test compound which binds to the polynucleotide is identified as a potential therapeutic agent for decreasing the activity of MASP-like serine protease.

13. A method of reducing the activity of MASP-like serine protease, comprising the steps of:

contacting a cell with a reagent which specifically binds to any polynucleotide of claim 1 or any MASP-like serine protease polypeptide of claim 4, whereby the activity of MASP-like serine protease is reduced.

14. A reagent that modulates the activity of a MASP-like serine protease polypeptide or a polynucleotide wherein said reagent is identified by the method of any of the claim 10 to 12.

15. A pharmaceutical composition, comprising:
the expression vector of claim 2 or the reagent of claim 14 and a pharmaceutically acceptable carrier.

- 71 -

16. Use of the pharmaceutical composition of claim 15 for modulating the activity of a MASP-like serine protease in extracellular matrix degradation.
- 5 17. Use of the pharmaceutical composition of claim 15 for defending against pathogens.
18. A cDNA encoding a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2.
- 10 19. The cDNA of claim 18 which comprises SEQ ID NO:1.
20. The cDNA of claim 18 which consists of SEQ ID NO:1.
- 15 21. An expression vector comprising a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2.
22. The expression vector of claim 21 wherein the polynucleotide consists of SEQ ID NO:1.
- 20 23. A host cell comprising an expression vector which encodes a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2.
24. The host cell of claim 23 wherein the polynucleotide consists of SEQ ID NO:1.
- 25 25. A purified polypeptide comprising the amino acid sequence shown in SEQ ID NO:2.
- 30 26. The purified polypeptide of claim 25 which consists of the amino acid sequence shown in SEQ ID NO:2.

27. A fusion protein comprising a polypeptide having the amino acid sequence shown in SEQ ID NO:2.
28. A method of producing a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2, comprising the steps of:
5 culturing a host cell comprising an expression vector which encodes the polypeptide under conditions whereby the polypeptide is expressed; and isolating the polypeptide.
- 10 29. The method of claim 28 wherein the expression vector comprises SEQ ID NO:1.
30. A method of detecting a coding sequence for a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2, comprising the steps of:
15 hybridizing a polynucleotide comprising 11 contiguous nucleotides of SEQ ID NO:1 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and detecting the hybridization complex.
- 20 31. The method of claim 30 further comprising the step of amplifying the nucleic acid material before the step of hybridizing.
32. A kit for detecting a coding sequence for a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2, comprising:
25 a polynucleotide comprising 11 contiguous nucleotides of SEQ ID NO:1; and instructions for the method of claim 30.
33. A method of detecting a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2, comprising the steps of:
30 contacting a biological sample with a reagent that specifically binds to the polypeptide to form a reagent-polypeptide complex; and

- 73 -

detecting the reagent-polypeptide complex.

34. The method of claim 33 wherein the reagent is an antibody.
- 5 35. A kit for detecting a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2, comprising:
an antibody which specifically binds to the polypeptide; and
instructions for the method of claim 33.
- 10 36. A method of screening for agents which can modulate the activity of a human MASP-like serine protease, comprising the steps of:
contacting a test compound with a polypeptide comprising an amino acid
sequence selected from the group consisting of: (1) amino acid sequences
which are at least about 50% identical to the amino acid sequence shown in
15 SEQ ID NO:2 and (2) the amino acid sequence shown in SEQ ID NO:2; and
detecting binding of the test compound to the polypeptide, wherein a test
compound which binds to the polypeptide is identified as a potential agent for
regulating activity of the human MASP-like serine protease.
- 20 37. The method of claim 36 wherein the step of contacting is in a cell.
38. The method of claim 36 wherein the cell is *in vitro*.
39. The method of claim 36 wherein the step of contacting is in a cell-free
25 system.
40. The method of claim 36 wherein the polypeptide comprises a detectable label.
41. The method of claim 36 wherein the test compound comprises a detectable
30 label.

- 74 -

42. The method of claim 36 wherein the test compound displaces a labeled ligand which is bound to the polypeptide.
43. The method of claim 36 wherein the polypeptide is bound to a solid support.
- 5 44. The method of claim 36 wherein the test compound is bound to a solid support.
- 10 45. A method of screening for agents which modulate an activity of a human MASP-like serine protease, comprising the steps of:
contacting a test compound with a polypeptide comprising an amino acid sequence selected from the group consisting of: (1) amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO:2 and (2) the amino acid sequence shown in SEQ ID NO:2; and
15 detecting an activity of the polypeptide, wherein a test compound which increases the activity of the polypeptide is identified as a potential agent for increasing the activity of the human MASP-like serine protease, and wherein a test compound which decreases the activity of the polypeptide is identified as a potential agent for decreasing the activity of the human MASP-like serine
20 protease.
46. The method of claim 45 wherein the step of contacting is in a cell.
47. The method of claim 45 wherein the cell is *in vitro*.
- 25 48. The method of claim 45 wherein the step of contacting is in a cell-free system.
- 30 49. A method of screening for agents which modulate an activity of a human MASP-like serine protease, comprising the steps of:

- 75 -

contacting a test compound with a product encoded by a polynucleotide which comprises the nucleotide sequence shown in SEQ ID NO:1; and detecting binding of the test compound to the product, wherein a test compound which binds to the product is identified as a potential agent for regulating the activity of the human MASP-like serine protease.

50. The method of claim 49 wherein the product is a polypeptide.

51. The method of claim 49 wherein the product is RNA.

52. A method of reducing activity of a human MASP-like serine protease, comprising the step of:

contacting a cell with a reagent which specifically binds to a product encoded by a polynucleotide comprising the nucleotide sequence shown in SEQ ID NO:1, whereby the activity of a human MASP-like serine protease is reduced.

53. The method of claim 52 wherein the product is a polypeptide.

54. The method of claim 53 wherein the reagent is an antibody.

55. The method of claim 52 wherein the product is RNA.

56. The method of claim 55 wherein the reagent is an antisense oligonucleotide.

57. The method of claim 56 wherein the reagent is a ribozyme.

58. The method of claim 52 wherein the cell is *in vitro*.

59. The method of claim 52 wherein the cell is *in vivo*.

60. A pharmaceutical composition, comprising:

- 76 -

a reagent which specifically binds to a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2; and
a pharmaceutically acceptable carrier.

- 5 61. The pharmaceutical composition of claim 60 wherein the reagent is an antibody.
62. A pharmaceutical composition, comprising:
a reagent which specifically binds to a product of a polynucleotide
10 comprising the nucleotide sequence shown in SEQ ID NO:1; and
a pharmaceutically acceptable carrier.
63. The pharmaceutical composition of claim 62 wherein the reagent is a ribozyme.
15
64. The pharmaceutical composition of claim 62 wherein the reagent is an antisense oligonucleotide.
65. The pharmaceutical composition of claim 62 wherein the reagent is an
20 antibody.
66. A pharmaceutical composition, comprising:
an expression vector encoding a polypeptide comprising the amino acid
sequence shown in SEQ ID NO:2; and
25 a pharmaceutically acceptable carrier.
67. The pharmaceutical composition of claim 66 wherein the expression vector comprises SEQ ID NO:1.

- 77 -

68. A method of treating a MASP-like serine protease disfunction related disease, wherein the disease is extracellular matrix degradation or a disease caused by pathogens, comprising the step of:
administering to a patient in need thereof a therapeutically effective dose of a
5 reagent that modulates a function of a human MASP-like serine protease, whereby symptoms of the MASP-like serine protease disfunction related disease are ameliorated.
69. The method of claim 68 wherein the reagent is identified by the method of
10 claim 36.
70. The method of claim 68 wherein the reagent is identified by the method of claim 45.
71. The method of claim 68 wherein the reagent is identified by the method of
15 claim 49.

Fig. 1

agtgtggtca gccctccgc tccctgccaa gcctggtcaa gaggatcatt gggggccgaa
atgctgagcc cggcctcttc ccgtggcagg ccctgatagt ggtggaggac acttcgagag
tgccaaatga caagtggttt gggagtgggg ccctgctctc tgcgtcctgg atcctcacag
cagctcatgt gctgcgtcc cagcgtagag acaccacggt gataccagtc tccaaggagc
atgtcaccgt ctacctgggc ttgcatgatg tgcgagacaa atcgggggca gtcaacagct
cagctgcccg agtggtgctc caccagact tcaacatcca aaactacaac cacgatatag
ctctgggtgca gctgcaggag cctgtgcccc tgggaccca cgttatgcct gtctgcctgc
caaggcttga gcctgaaggc ccggccccc acatgctggg cctggtggcc ggctggggca
tctccaatcc caatgtgaca gtggatgaga tcatcagcag tggcacacgg accttgctag
atgtcctaca gtatgtcaag ttaccctggg tgctcagc tgagtgcaa actagctatg
agtcccgtc gggcaattac agcgtcacgg agaacatggt ctgtgctggc tactacgagg
gcggcaaaga cacgtgcctt ggagatagcg gtggggcctt tgcaccttt gatgacttga
gccagcgctg ggtggtgcaa ggcctggtgt cctggggggg acctgaagaa tgcggcagca
agcaggtcta tggagtctac acaaaggctt ccaattacgt ggactgggtg tgggagcaga
tg

Fig. 2

CGQPSRSLPS LVKRIIGRN AEPGLFPWQA LIVVEDTSRV PNDKWFGSGA LLSASWILTA
AHVLSRQRRD TTVIPVSKEH VTVYLGLHDV RDKSGAVNSS AARVVLHPDF NIQNYNHDA
LVQLQEPVPL GPHVMPVCLP RLEPEGPAPH MLGLVAGWGI SNPNVTVDEI ISSGTRTLSD
VLQYVKLPVV PHAECKTSYE SRSGNYSVTE NMFCAGYYEG GKDTCLGDSG GAFVIFDDL
QRWVVQGLVS WGGPEECGSK QVYGVYTKVS NYVDWVWEQM

Fig. 3

AB009074

MGCFWLFVLSFFTPAGANVTELTGMFGQIQSPNYPEGYPDDLDTWNITVPAGFRIKLYF
MHFDLEPSYLCEYDYVKVVSDEDLLETFCGRESTDTEQIPGREFVLTSNTMAVTLKTDF
SNEERFTGFRAHYSAVDIDECVEKLDEALVCDHFCHNYIGGYCSCRFSYILHTDNRTCK
VECNDNVFTEKSGVIRSPDYPNPYPKSSDCLYRIELEEFGAISLEFADMFDLEDHDPDVPC
PYDYLIKAGSMAFGPPFCGDKSPGKIDTNSSSVQILFHSGGSGDNQGWKMSFLSVGTPCP
ILHPPANGTIEPVQDQYSFKDQVQIRCNTGFKFVKAGDEVEAYQIECQKSGSWSSSIPTC
KVIDCGAPRPLEDGIITFHSKENLTLFRSHIEYSCPALHYHLEPELN'TTYTCMENGFWVN
DELGTDLP TCQPVCGRPAPPLPPHNKRIIGGRTAAPGFFPWQLLIVEDVSRVPKDKWFG
SGALLSRTWVLTAHVLR SQRRDTITLVPSEYVTIYLG LHDVRQKEAAAKRTVEKIILHK
AFDPRTYNNDIALVKMKDKVSMNVFVMPLCLPSLHQEME EPQPN TLGLVAGWGITNPNLT
LDDASGSDQATLSN ILQYVKLPVTLQAECKSSYESRSDSYNVTDNMFCAGFYEGGKDTCL
GDSGGAFTYDSSTQSWVAQGLVSWGGPEKCGSKRVYGVYTKISKYARWLADKMSNSSD

Fig. 4

IvGGreaqpgsfgsPwqvslqvrsgggsrkhfCGGsLisenwVLTAAHCvsgaa
sapassvrVSlsvrlGehnlsltegteqkfdvkktiivHpnyndpdtldngaYdnDiA
LlkLkspgvtlgdtvrpicLpsassdlpvGttctvsGwGrrptknlg
lsdtLqevvvpvsretCrsayeyggtDkvefvtdnmiCagalggkdaCqGDSGG
PLvcsdgnrdgrwelvGivSwGsygCargnkPGvytrVssyldWI

FIG. 5 BLASTP - Query = 142_2_TR3 (SEQ ID NO:2); Hit =
 trembl|AB009074|AB009074_1 (SEQ ID NO:3)

This hit is scoring at : 5e-108 (expectation value)

Alignment length (overlap) : 282

Identities : 64 %

Scoring matrix : BLOSUM62 (used to infer consensus pattern)

Database searched : nrdb

Q: 1 CGQPSRSLPSLVKRIIGGRNAEPGLFPWQALIVVEDTSRVPNDKWFSGALLSASWILTA
 CG:P:R:LP. KRIIGGR.A.PG.FPWQ.LIVVED.SRVP.DKWFGSGALLS.:W:LTA
 H: 434 CGRPAPPLPPHNKRIIGGRTAAPGFFPWQLLIVVEDVSRVPKDKWFSGALLSRTWVLTA

AHVLRSQRRDTTVIPVSKEHVTVYLGLHDVRDKSGAVNSSAARVVLHDPDFNIQYNHNDIA
 AHVLRSQRRDT... V..E:VT:YLGLHDVR.K..A.....:LH. F: ..YN:DIA
 AHVLRSQRRDTITL-VPSEYVTIYLGLHDVRQKEAAAKRTVEKIILHKAFDPRTYNNDIA

active_HIS

Active_ASP

LVQLQEPVPLGPHVMPVCLPRL--EPEGPAHMLGLVAGWGISNPNVTVDIISGTRTL
 LV::::V...VMP:CLP.L E E P.P.:LGLVAGWGI:NPN:T:D:.S...TL
 LVKMKDKVSMNVFVMPCLCLPSLHQEMEEPQPNTLGLVAGWGITNPNLTLDASGSDQATL

SDVLQYVKLPVVPHAECKTSYESRSGNYSVTENMFCAGYYEGGKDTCLGDSGGAFVIFDD
 S::LQYVKLPV. .AECK:SYESRS.:Y:VT:NMFAG:YEGGKDTCLGDSGGAF::D.
 SNILQYVKLPVTLQAECKSSYESRSDSYNVTDNMFCAGFYEGGKDTCLGDSGGAFITYDS

TRYPSIN_SER

LSQRWVVGGLVSWGGPEECGSKQVYGVYTKVSNYVDWVWEQM 280
 :Q.WV.QGLVSWGGPE:CGSK:VYGVYTK:S.Y. W: ::M
 STQSWVAQGLVSWGGPEKCGSKRVYGVYTKISKYARWLADKM 714

BLOCKS

FIG. 6

Prosites search results

Access#	From->To	Name	Doc#
PS00135	223->235	TRYPSIN_SER	PDOC00124

FIG. 7

BLOCKS search results

AC#	Description	Strength	Score
BL00134A	Serine proteases, trypsin family, histidine p	1500	1320
AA#	46 GSGALLSASWILTAAHV (SEQ ID NO:5)		
BL00134C	Serine proteases, trypsin family, histidine p	1245	1271
AA#	262 YGVYTKVSNYVDWV (SEQ ID NO:6)		
BL00495O	Apple domain proteins.	1756	1245
AA#	253 pEeCGsKqvyGVYTKVSnYVDWvwEqm (SEQ ID NO:7)		
BL00495N	Apple domain proteins.	1945	1215
AA#	214 AGYyEGGKDtClGDSGGafVifddlsqRwVvqglv (SEQ ID NO:8)		

FIG. 8

HMMPFAM - alignment of 142_2_TR3 (SEQ ID NO:2) against pfam|hmm|trypsin
SEQ ID NO:4)

Trypsin

This hit is scoring at : 244.0; Expect = 1.1-76

Scoring matrix : BLOSUM62 (used to infer consensus pattern)

Q: 15 IIGGRNAEPGLF--PWQALIVVEDTSRvpNDKWFGSGALLSASWILTAHVLRs-Qrrdt
I:GGR.A:PG F PWQ. : V... : K F .G:L:S.:W:LTAAH....
H: 1 IvGGreaqpgsfgsPwqvslqvrsggg..srkhfCGGsLisenwVLTAAHCVsgaa....

tvIPVSKEHVTv----YLGLHDVRDKSGA-VNSSAAR-VVLHPDFNIQNYN-----HDIA
.....V.V LG H:... ..G. : ::HP::N :DIA
...sapassvrVSlsvrlGehnlsltegteqkfdvkktiivHpnynpdtldngaYdnDiA

LVQLQEP-VPLGPHVMPVCLPRlepEGP--APHMLGLVAGWGI--SNPNVtvdeiissgt
L::L::P V.LG. V.P:CLP.V:GWG :.
LlkLkspgvtlgtvvpicLps...assdlpvGttctvsGwGrrptknlg.....

rtLSDVLQYVKLPVVPPhAECKTSYEsRSGN---YSVTENMFCAGYYeGGKDTCLGDSGG
LSD.LQ V :PVV....C:::YE .G. VT:NM.CAG . GGKD.C GDSGG
..lsdtLqevvvpvvsretCrsaye.yggtdDkvefvtdnmiCagal.ggkdaCqGDSGG

AFVifdDLSQ----RWVVQGLVSWGGpEECGSKQVYGvYTKVSNYVDWV 276
..V . RW : G:VSWG. C.. . GVYT:VS:Y:DW:
PLv...csdgnrdgrwelvGivSwGs.ygCargnkPGvytrVssyldWI 259

SEQUENCE LISTING

<110> Bayer AG

<120> REGULATION OF HUMAN MASP-LIKE SERINE PROTEASE

<130> Lio 116

<150> USSN 60/221,211

<151> 2000-07-27

<160> 4

<170> PatentIn version 3.1

<210> 1

<211> 842

<212> DNA

<213> Homo sapiens

<400> 1

```

agtgtggtca gccctcccgc tcctgccaa gcctggtcaa gaggatcatt gggggccgaa      60
atgctgagcc cggcctcttc ccgtggcagg ccctgatagt ggtggaggac acttcgagag      120
tgccaaatga caagtggttt gggagtgggg ccctgctctc tgcgtcctgg atcctcacag      180
cagctcatgt gctgcgctcc cagcgtagag acaccacggt gataccagtc tccaaggagc      240
atgtcacctg ctacctgggc ttgcatgatg tgcgagacaa atcgggggca gtcaacagct      300
cagctgcccc agtgggtgct caccagact tcaacatcca aaactacaac cacgatatag      360
ctctggtgca gctgcaggag cctgtgcccc tgggaccca cgttatgcct gtctgcctgc      420
caaggcttga gcctgaaggc ccggcccccc acatgctggg cctggtggcc ggctggggca      480
tctccaatcc caatgtgaca gtggatgaga tcatcagcag tggcacacgg accttgtcag      540
atgtcctaca gtatgtcaag ttaccctggg tgccctcacgc tgagtgcaaa actagctatg      600
agtcccgtc gggcaattac agcgtcacgg agaacatggt ctgtgctggc tactacgagg      660
gcggcaaaga cacgtgcctt ggagatagcg gtggggcctt tgtcatcttt gatgacttga      720
gccagcgctg ggtggtgcaa ggcctggtgt cctggggggg acctgaagaa tgcggcagca      780
agcaggtcta tggagtctac acaaaggtct ccaattacgt ggactgggtg tgggagcaga      840
tg                                                                 842

```

<210> 2

<211> 280

<212> PRT

<213> Homo sapiens

<400> 2

Cys Gly Gln Pro Ser Arg Ser Leu Pro Ser Leu Val Lys Arg Ile Ile
 1 5 10 15

Gly Gly Arg Asn Ala Glu Pro Gly Leu Phe Pro Trp Gln Ala Leu Ile
 20 25 30

Val Val Glu Asp Thr Ser Arg Val Pro Asn Asp Lys Trp Phe Gly Ser
 35 40 45

Gly Ala Leu Leu Ser Ala Ser Trp Ile Leu Thr Ala Ala His Val Leu
 50 55 60

Arg Ser Gln Arg Arg Asp Thr Thr Val Ile Pro Val Ser Lys Glu His
 65 70 75 80

Val Thr Val Tyr Leu Gly Leu His Asp Val Arg Asp Lys Ser Gly Ala
 85 90 95

Val Asn Ser Ser Ala Ala Arg Val Val Leu His Pro Asp Phe Asn Ile
 100 105 110

Gln Asn Tyr Asn His Asp Ile Ala Leu Val Gln Leu Gln Glu Pro Val
 115 120 125

Pro Leu Gly Pro His Val Met Pro Val Cys Leu Pro Arg Leu Glu Pro
 130 135 140

Glu Gly Pro Ala Pro His Met Leu Gly Leu Val Ala Gly Trp Gly Ile
 145 150 155 160

Ser Asn Pro Asn Val Thr Val Asp Glu Ile Ile Ser Ser Gly Thr Arg
 165 170 175

Thr Leu Ser Asp Val Leu Gln Tyr Val Lys Leu Pro Val Val Pro His
 180 185 190

Ala Glu Cys Lys Thr Ser Tyr Glu Ser Arg Ser Gly Asn Tyr Ser Val
 195 200 205

Thr Glu Asn Met Phe Cys Ala Gly Tyr Tyr Glu Gly Gly Lys Asp Thr
 210 215 220

Cys Leu Gly Asp Ser Gly Gly Ala Phe Val Ile Phe Asp Asp Leu Ser
 225 230 235 240

Gln Arg Trp Val Val Gln Gly Leu Val Ser Trp Gly Gly Pro Glu Glu
 245 250 255

Cys Gly Ser Lys Gln Val Tyr Gly Val Tyr Thr Lys Val Ser Asn Tyr
 260 265 270

Val Asp Trp Val Trp Glu Gln Met
 275 280

<210> 3
 <211> 719
 <212> PRT
 <213> Homo sapiens

<400> 3

Met Gly Cys Phe Trp Leu Phe Val Leu Ser Phe Phe Thr Pro Ala Gly
 1 5 10 15

Ala Asn Val Thr Glu Leu Thr Gly Met Phe Gly Gln Ile Gln Ser Pro
 20 25 30

Asn Tyr Pro Glu Gly Tyr Pro Ser Asp Leu Asp Val Thr Trp Asn Ile
 35 40 45

Thr Val Pro Ala Gly Phe Arg Ile Lys Leu Tyr Phe Met His Phe Asp
 50 55 60

Leu Glu Pro Ser Tyr Leu Cys Glu Tyr Asp Tyr Val Lys Val Val Ser
 65 70 75 80

Asp Glu Asp Leu Leu Glu Thr Phe Cys Gly Arg Glu Ser Thr Asp Thr
 85 90 95

Glu Gln Ile Pro Gly Arg Glu Phe Val Leu Thr Asn Ser Asn Thr Met
 100 105 110

Ala Val Thr Leu Lys Thr Asp Phe Ser Asn Glu Glu Arg Phe Thr Gly
 115 120 125

Phe Glu Ala His Tyr Ser Ala Val Asp Ile Asp Glu Cys Val Glu Lys

130				135				140							
Leu 145	Asp	Glu	Ala	Leu	Val 150	Cys	Asp	His	Phe	Cys 155	His	Asn	Tyr	Ile	Gly 160
Gly	Tyr	Tyr	Cys	Ser 165	Cys	Arg	Phe	Ser	Tyr 170	Ile	Leu	His	Thr	Asp 175	Asn
Arg	Thr	Cys	Lys 180	Val	Glu	Cys	Asn	Asp 185	Asn	Val	Phe	Thr	Glu 190	Lys	Ser
Gly	Val	Ile 195	Arg	Ser	Pro	Asp	Tyr 200	Pro	Asn	Pro	Tyr	Pro 205	Lys	Ser	Ser
Asp	Cys 210	Leu	Tyr	Arg	Ile	Glu 215	Leu	Glu	Glu	Gly	Phe 220	Ala	Ile	Ser	Leu
Glu 225	Phe	Ala	Asp	Met	Phe 230	Asp	Leu	Glu	Asp	His 235	Pro	Asp	Val	Pro	Cys 240
Pro	Tyr	Asp	Tyr	Leu 245	Lys	Ile	Lys	Ala	Gly 250	Ser	Met	Ala	Phe	Gly 255	Pro
Phe	Cys	Gly	Asp 260	Lys	Ser	Pro	Gly	Lys 265	Ile	Asp	Thr	Asn	Ser 270	Ser	Ser
Val	Gln	Ile 275	Leu	Phe	His	Ser	Gly 280	Gly	Ser	Gly	Asp	Asn 285	Gln	Gly	Trp
Lys	Met 290	Ser	Phe	Leu	Ser	Val 295	Gly	Thr	Pro	Cys	Pro 300	Ile	Leu	His	Pro
Pro 305	Ala	Asn	Gly	Thr	Ile 310	Glu	Pro	Val	Gln	Asp 315	Gln	Tyr	Ser	Phe	Lys 320
Asp	Gln	Val	Gln	Ile 325	Arg	Cys	Asn	Thr	Gly 330	Phe	Lys	Phe	Val	Lys 335	Ala
Gly	Asp	Glu	Val 340	Glu	Ala	Tyr	Gln	Ile 345	Glu	Cys	Gln	Lys	Ser 350	Gly	Ser
Trp	Ser	Ser 355	Ser	Ile	Pro	Thr	Cys 360	Lys	Val	Ile	Asp	Cys 365	Gly	Ala	Pro

Arg Pro Leu Glu Asp Gly Ile Ile Thr Phe His Ser Lys Glu Asn Leu
 370 375 380

Thr Leu Phe Arg Ser His Ile Glu Tyr Ser Cys Pro Ala Leu His Tyr
 385 390 395 400

His Leu Glu Pro Glu Leu Asn Thr Thr Tyr Thr Cys Met Glu Asn Gly
 405 410 415

Phe Trp Val Asn Asp Glu Leu Gly Thr Asp Leu Pro Thr Cys Gln Pro
 420 425 430

Val Cys Gly Arg Pro Ala Arg Pro Leu Pro Pro His Asn Lys Arg Ile
 435 440 445

Ile Gly Gly Arg Thr Ala Ala Pro Gly Phe Phe Pro Trp Gln Leu Leu
 450 455 460

Ile Val Val Glu Asp Val Ser Arg Val Pro Lys Asp Lys Trp Phe Gly
 465 470 475 480

Ser Gly Ala Leu Leu Ser Arg Thr Trp Val Leu Thr Ala Ala His Val
 485 490 495

Leu Arg Ser Gln Arg Arg Asp Thr Ile Thr Leu Val Pro Ser Glu Tyr
 500 505 510

Val Thr Ile Tyr Leu Gly Leu His Asp Val Arg Gln Lys Glu Ala Ala
 515 520 525

Ala Lys Arg Thr Val Glu Lys Ile Ile Leu His Lys Ala Phe Asp Pro
 530 535 540

Arg Thr Tyr Asn Asn Asp Ile Ala Leu Val Lys Met Lys Asp Lys Val
 545 550 555 560

Ser Met Asn Val Phe Val Met Pro Leu Cys Leu Pro Ser Leu His Gln
 565 570 575

Glu Met Glu Glu Pro Gln Pro Asn Thr Leu Gly Leu Val Ala Gly Trp
 580 585 590

Gly Ile Thr Asn Pro Asn Leu Thr Leu Asp Asp Ala Ser Gly Ser Asp
 595 600 605

Gln Ala Thr Leu Ser Asn Ile Leu Gln Tyr Val Lys Leu Pro Val Thr
 610 615 620

Leu Gln Ala Glu Cys Lys Ser Ser Tyr Glu Ser Arg Ser Asp Ser Tyr
 625 630 635 640

Asn Val Thr Asp Asn Met Phe Cys Ala Gly Phe Tyr Glu Gly Gly Lys
 645 650 655

Asp Thr Cys Leu Gly Asp Ser Gly Gly Ala Phe Ile Thr Tyr Asp Ser
 660 665 670

Ser Thr Gln Ser Trp Val Ala Gln Gly Leu Val Ser Trp Gly Gly Pro
 675 680 685

Glu Lys Cys Gly Ser Lys Arg Val Tyr Gly Val Tyr Thr Lys Ile Ser
 690 695 700

Lys Tyr Ala Arg Trp Leu Ala Asp Lys Met Ser Asn Ser Ser Asp
 705 710 715

<210> 4
 <211> 259
 <212> PRT
 <213> Homo sapiens

<400> 4

Ile Val Gly Gly Arg Glu Ala Gln Pro Gly Ser Phe Gly Ser Pro Trp
 1 5 10 15

Gln Val Ser Leu Gln Val Arg Ser Gly Gly Gly Ser Arg Lys His Phe
 20 25 30

Cys Gly Gly Ser Leu Ile Ser Glu Asn Trp Val Leu Thr Ala Ala His
 35 40 45

Cys Val Ser Gly Ala Ala Ser Ala Pro Ala Ser Ser Val Arg Val Ser
 50 55 60

Leu Ser Val Arg Leu Gly Glu His Asn Leu Ser Leu Thr Glu Gly Thr

65		70		75		80
Glu Gln Lys Phe Asp Val Lys Lys Thr Ile Ile Val His Pro Asn Tyr						
	85			90		95
Asn Pro Asp Thr Leu Asp Asn Gly Ala Tyr Asp Asn Asp Ile Ala Leu						
	100			105		110
Leu Lys Leu Lys Ser Pro Gly Val Thr Leu Gly Asp Thr Val Arg Pro						
	115			120		125
Ile Cys Leu Pro Ser Ala Ser Ser Asp Leu Pro Val Gly Thr Thr Cys						
	130			135		140
Thr Val Ser Gly Trp Gly Arg Arg Pro Thr Lys Asn Leu Gly Leu Ser						
	145			150		155
						160
Asp Thr Leu Gln Glu Val Val Val Pro Val Val Ser Arg Glu Thr Cys						
				165		170
						175
Arg Ser Ala Tyr Glu Tyr Gly Gly Thr Asp Asp Lys Val Glu Phe Val						
	180			185		190
Thr Asp Asn Met Ile Cys Ala Gly Ala Leu Gly Gly Lys Asp Ala Cys						
	195			200		205
Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Asp Gly Asn Arg Asp						
	210			215		220
Gly Arg Trp Glu Leu Val Gly Ile Val Ser Trp Gly Ser Tyr Gly Cys						
	225			230		235
						240
Ala Arg Gly Asn Lys Pro Gly Val Tyr Thr Arg Val Ser Ser Tyr Leu						
	245			250		255
Asp Trp Ile						

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10404 A3

(51) International Patent Classification⁷: **C12N 15/57**,
15/62, 5/10, 9/64, C07K 16/40, C12Q 1/68, 1/37, G01N
33/68, 33/577, A61K 31/711, 31/713, 38/43, 39/395

(21) International Application Number: PCT/EP01/08181

(22) International Filing Date: 16 July 2001 (16.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,211 27 July 2000 (27.07.2000) US

(71) Applicant (for all designated States except US): **BAYER
AKTIENGESELLSCHAFT** [DE/DE]; 51368 Lev-
erkusen (DE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **XIAO, Yonghong**
[US/US]; 78 Dana Street #1, Cambridge, MA 02138 (US).

(74) Common Representative: **BAYER AKTIENGE-
SELLSCHAFT**; 51368 Leverkusen (DE).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(88) Date of publication of the international search report:
13 June 2002

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.



WO 02/10404 A3

(54) Title: REGULATION OF HUMAN MASP-LIKE SERINE PROTEASE

(57) Abstract: Reagents which regulate human MASP-like serine protease activity and reagents which bind to human MASP-like serine protease gene products can be used to regulate extracellular matrix degradation. Such regulation is particularly useful for defending against pathogens, such as viruses, bacteria, mycoplasma, fungi, protozoa, helminths, rikettsia, chloamydiae, parasites, prions, and the like.

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/EP 01/08181

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/57 C12N15/62 C12N5/10 C12N9/64 C07K16/40
C12Q1/68 C12Q1/37 G01N33/68 G01N33/577 A61K31/711
A61K31/713 A61K38/43 A61K39/395

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, CAB Data, SEQUENCE SEARCH, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>WO 01 38501 A (ZYMOGENETICS INC) 31 May 2001 (2001-05-31) the whole document -& DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 22 June 2001 (2001-06-22) J.L. HOLLOWAY: "Sequence 1 and 2 from patent Wo0138501" XP002195187 EMBL Accession nos. AX150967 and AX150968; abstract</p> <p style="text-align: center;">--- -/--</p>	1-71



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

Z document member of the same patent family

Date of the actual completion of the international search

5 April 2002

Date of mailing of the international search report

17/04/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040. Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 01/08181

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>WO 01 40451 A (THIEL STEFFEN ; JENSENIUS JENS CHRISTIAN (DK)) 7 June 2001 (2001-06-07) the whole document -& DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 22 June 2001 (2001-06-22) J.C. JENSENIUS AND S. THIEL: "Sequences 1 and 2 from patent W00140451" XP002195188 EMBL Sequence Accession nos. AX156466 and AX156467; abstract</p> <p>---</p>	1-71
E	<p>WO 01 71004 A (INCYTE GENOMICS INC ; PATTERSON CHANDRA (US); TRIBOULEY CATHERINE M) 27 September 2001 (2001-09-27) the whole document -& DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 10 October 2001 (2001-10-10) H. YUE ET AL.: "Sequences 1 and 2 from patent W00171004 " XP002195189 EMBL Sequence Accession nos. AX254412 and AX254423; abstract</p> <p>---</p>	1-71
X	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! Hinxton, UK; 23 May 2000 (2000-05-23) D.M. MUZNY ET AL.: "Homo sapiens 3 BAC RP11-460F1 (Roswell Park Cancer Institute HUMAN BAC Library) complete Sequence" XP002195190 EMBL Accession no. AC069069; 100% identity (100% ungapped) in 842nt overlap; abstract</p> <p>---</p>	1
T	<p>DAHL MADRS ET AL: "MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway." IMMUNITY, vol. 15, no. 1, July 2001 (2001-07), pages 127-135, XP002195186 ISSN: 1074-7613 the whole document</p> <p>-----</p>	

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: (13-17,52-55,60,62,68-71)-partially

Present claims (13-17,52-55,59,60,62,68-71)-partially relate to a method of modulating the activity of a MASP-like serine protease or a MASP-like serine protease expressing cell, respectively a method of treating or preventing a disorder by using an agent that modulates the activity or expression of said polypeptide without giving a true technical characterization. Moreover no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to antibodies, antisense molecules, ribozymes, triple helix molecules, polypeptides and nucleic acids, the structure of which can be directly derived from SEQ ID Nos. 1 and 2.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 01/08181

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
WO 0138501	A	31-05-2001	AU	1624501 A	04-06-2001
			WO	0138501 A2	31-05-2001
WO 0140451	A	07-06-2001	AU	1693801 A	12-06-2001
			WO	0140451 A2	07-06-2001
WO 0171004	A	27-09-2001	AU	4748301 A	03-10-2001
			WO	0171004 A2	27-09-2001